C. H. Ngai and Z. Tian

Sound symbolism through the ages: A longitudinal analysis of vowel and tonal Patterns in Chinese names across generations

Introduction

With renewed interest in the sound symbolism of names, a growing number of studies could be found examining systematic sound-meaning mapping in various languages (e.g., English (Pitcher et al., 2013), French (Suire et al., 2019), and Mandarin Chinese (Ngai & Kilpatrick, 2023), and many more). Systematically, these studies have found that male names are more likely to contain low/back vowels, while female names tend to incorporate high/front vowels. However, except Suire et al. (2019), most existing work has mostly focused on synchronic analysis, providing little insight into how naming patterns may evolve diachronically. If such sound-meaning mappings are truly rooted in sexual selection (Pitcher et al., 2013; Suire et al., 2019) or the Frequency codes hypothesis[1] (Ohala, 1984, 1994), such patterns should persist across generations of speakers. To address this issue, this study analyses whether sound-gender patterns previously found in Mandarin (low/back vowel and tone 3 in male names; high/front vowels in female names (Ngai & Kilpatrick, 2023)) persist across time while accounting for regional differences.

Methods and Materials

This study examines the sound-gender mapping in Mandarin Chinese names using an online name database from the Shanghai police[2]. Names were first transcribed phonemically and filtered by year of birth (1947-2016) to ensure sizeable samples across years. Table 1 illustrates the number of male and female names per year in the current dataset. Names that could not be transcribed into IPA and Pinyin were also discarded. The final dataset consisted of a total of 200,392 names (106, 436 male and 93,956 female, or 1,520 male and 1342 female per year). Vowel (height and frontness) and tonal features (tones 1 and 3) were extracted from transcriptions.

Following Suire et al. (2019), linear mixed effect models built with each phonetic feature (e.g., height, frontness, tone 1, and tone 3) as the outcome, and year of birth, gender, and their interactions as fixed predictors, including random intercepts for the provinces of birth. Model comparison via ANOVA was conducted to determine if gender and birth year interaction significantly improved fit, indicating divergent or convergent patterns along gender and temporal lines.

Results

Only coefficients relating to the interaction between gender and birth year are discussed below. The results of linear mixed effect models demonstrated significant changes in the frequency of front vowels ([i], [y], and [a]) and tone 3 frequencies over time that differ between male and female names. Results from the front vowel model found a highly significant interaction between genders and birth year ($\beta = 0.186$, $SE = 0.032$, $t = 5.881$, $p < .001$), suggesting a widening gap in front vowels adaptation between genders over time. No significant interaction between gender and birth year was observed for vowel height. As for tones, only results from tone 3 discovered a significant interaction between gender and birth year ($\beta = 0.368$, $SE = 0.021$, $t = 17.299$, $p < .001$), indicating a narrowing in the frequency of tone 3 between male and female names as time progresses.

---

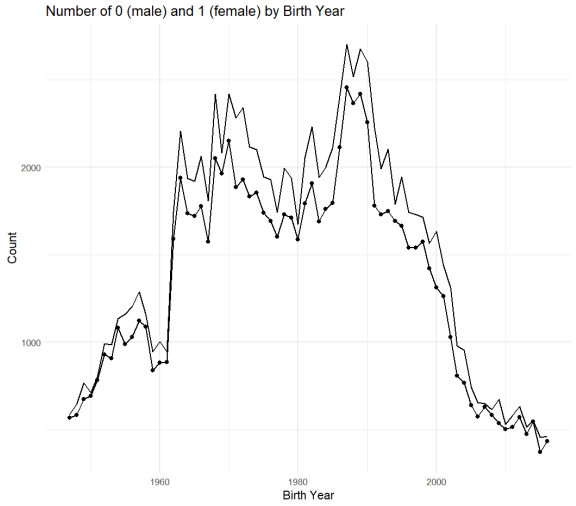
[1] The Frequency Codes Hypothesis posits that the observed sound-gender associations in naming practices stem from sound-to-size mappings, whereby female names exhibit a preference for phonemes with acoustically higher frequencies (associated with smallness), while male names favor phonemes with acoustically lower frequencies (associated with largeness).

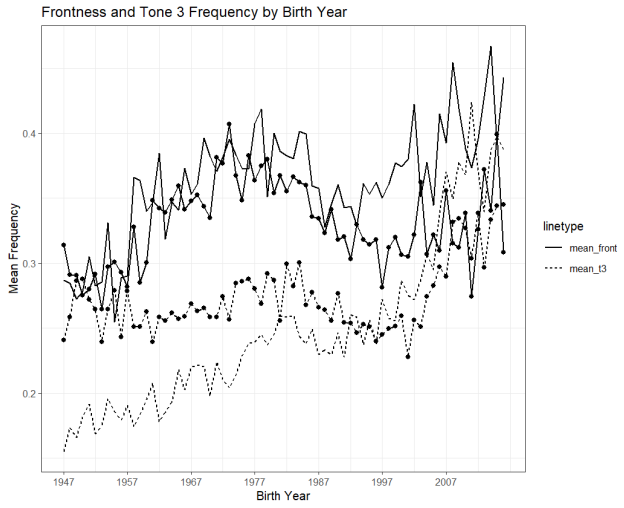
[2] https://breached.to/Thread-2022-SHGA-SHanghai-Gov-National-Police-750K-Sample

Discussion

The current study provided insights into the diachronic changes in sound-gender mappings in Mandarin Chinese names. Specifically, our analysis revealed a significant interaction between gender and birth year in front vowels and tone 3 frequencies, indicating evolving sound-gender association over time. Interestingly, patterns observed for front vowels differ from those in tone 3 frequencies (see Table 2). While the gap in front vowels between genders widens, a narrowing difference between genders was observed in tone 3. Results pertaining to front vowels demonstrated that mappings found in synchronic studies (e.g., Ngai & Kilpatrick, 2023) could also be found in a diachronic study. It is worth noting that our study incorporated regional differences by including random intercepts for the provinces of birth. This approach accounts for potential influence from varying first languages, ensuring the robustness of our analysis. Combined, these results demonstrated both persistence and evolution in sound symbolic sound-meaning mappings, highlighting the sound symbolism sound meaning could also be influenced by culture.

Table 1. (left) Number of male (dotted) and female (line) names in the database.
Table 2. (right) Frequency of front vowels and tone 3 in male (dotted) and female (line) names overtime.





References

Ngai, C. H., & Kilpatrick, A. (2023). Deriving gender from phonemes: Evidence from Mandarin names using the Random Forest algorithm. *45th DGfS AG15: Ideophones and Lexicalized Iconicity in Language*.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica*, *41*(1), 1–16.

Ohala, J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In *Sound Symbolism* (pp. 325–347). Cambridge University Press.

Pitcher, B. J., Mesoudi, A., & McElligott, A. G. (2013). Sex-biased sound symbolism in English-language first names. *PLoS ONE*, *8*(6), e64825.

Suire, A., Mesa, A. B., Raymond, M., & Barkat-Defradas, M. (2019). Sex-biased sound symbolism in French first names. *Evolutionary Human Sciences*, *1*(e7), 1–17.