



Indexing political identity through syntactic variables: Classifier specificity in pro-Taiwan and pro-China Taiwan Mandarin speakers

Po-Hsuan Huang & Stephanie Shih
{pohsuan, shihs}@usc.edu

University of Southern California



Political indexicality

- Linguistic variables have been found to **index political identity**:
 - Hall-Lew *et al.* (2010; 2012) found that:
 - $Ir[a:]q \sim Ir[\ae]q$ could index political liberalism ~ conservatism.
 - Republicans are significantly more likely than Democrats to use /æ/ in the House of Representatives.
 - Sloman *et al.* (2021) extracted words more likely to be used by Republican/Democrat politicians, and found that speakers are able to guess which word is more likely to be used by Republicans/Democrats at an above-chance rate.



Political indexicality

- Linguistic variables have been found to **index political identity**.
- Previous studies, however, focused on **phonetic and lexical** variables.
- We investigated whether nuanced morphosyntactic distinctions may also serve to index political identity:
 - **Classifier specificity among pro-Taiwan & pro-China Taiwan Mandarin speakers.**



Classifier specificity

- In Mandarin, classifiers are paired w/ the noun referents (Hsieh, 2009; Her & Hsieh, 2010; Her & Lai, 2012; Yip

& Rimmington, 2015).

- Nouns may only go w/ certain classifiers.

一 隻 狗。
Yi zhi gou.
one CL:animal dog
“A dog.”

一 個 狗。
Yi ge gou.
one CL:general dog
“A dog.”

*一 顆 狗。
Yi ke gou.
one CL:round dog
Intended: “A dog.”



Classifier specificity

- In Mandarin, classifiers are paired w/ the noun referents (Hsieh, 2009; Her & Hsieh, 2010; Her & Lai, 2012; Yip

& Rimmington, 2015).

- Nouns may only go w/ certain classifiers.

- Nouns can go w/ more than one classifier:

- An array of classifiers to choose from.

→ The “general” classifier **ge 個** vs.
specific classifiers

一 隻 狗。
Yi zhi gou.
one CL:animal dog
“A dog.”

一 個 狗。
Yi ge gou.
one CL:general dog
“A dog.”

*一 顆 狗。
Yi ke gou.
one CL:round dog
Intended: “A dog.”

Classifier specificity & sociolinguistic meaning

- Zheng & Liu (2023) found that classifier specificity indexed different perceived personal traits among Mandarin speakers recruited on a social media platform:
 - The use of the specific classifiers was rated as more meticulous & steady.

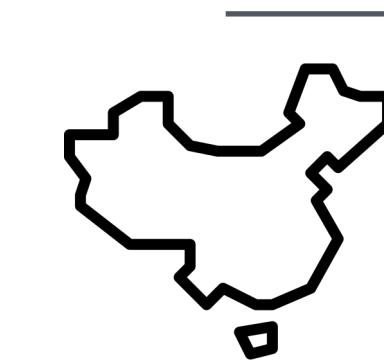
Classifier specificity as political indexicality

- Taiwanese political polarization is based less on a left-right ideology but more on the *pro-Taiwan* vs. *pro-China* division (Hsiao & Cheng, 2014).
- In this study, we investigate: **(Q1) Can classifier specificity likewise be used to index political identity among pro-Taiwan & pro-China Taiwan Mandarin speakers?**



Classifier specificity as political indexicality

- Hall-Lew & van Eynhoven (2025) propose that political indexicality is typically derived from *politicizing* an existing indexicality.
- Intuitively, the indexicality of variety (Taiwan Mandarin ~ Chinese Mandarin) is the most likely target. → **(Q2) Would the variety differences b/t Taiwan Mandarin & Chinese Mandarin show a similar directionality?**



Chinese Mandarin



Taiwan Mandarin

pro-China

pro-Taiwan



Data

- To answer these questions, we collected spontaneous speech data from 2,841 publicly-accessible YouTube videos created by Taiwanese & Chinese creators (2,345,967 sentences; 11,453.56 hours of speech).



A screenshot from one of the videos, HahaTai 哈哈台 (2024, Oct. 24). A person is being interviewed.



Methods

Classifier specificity

- Classifier-noun pairs (adjacent & nonadjacent) were extracted using Stanford CoreNLP (212,150 pairs extracted).
- We modeled classifier specificity on two levels:
 - Continuous specificity: the mean absolute **pointwise mutual information** (i.e., the degree of association) b/t a classifier & all the co-occurring nouns
 - Binary distinction: **general** (ge 個 & jian 件) **vs.** **specific** (others) **classifiers** (based on their PMI distribution)

$$PMI(classifier, n) = \frac{P_{corpus}(n | classifier)}{P_{corpus}(n)}$$



Methods

Social factor labeling

- Political identity: labeled based on the content creators/news media.



A pro-China legislator criticizing Trump's tariff policies (TVBS Talk, 2025, April. 22).



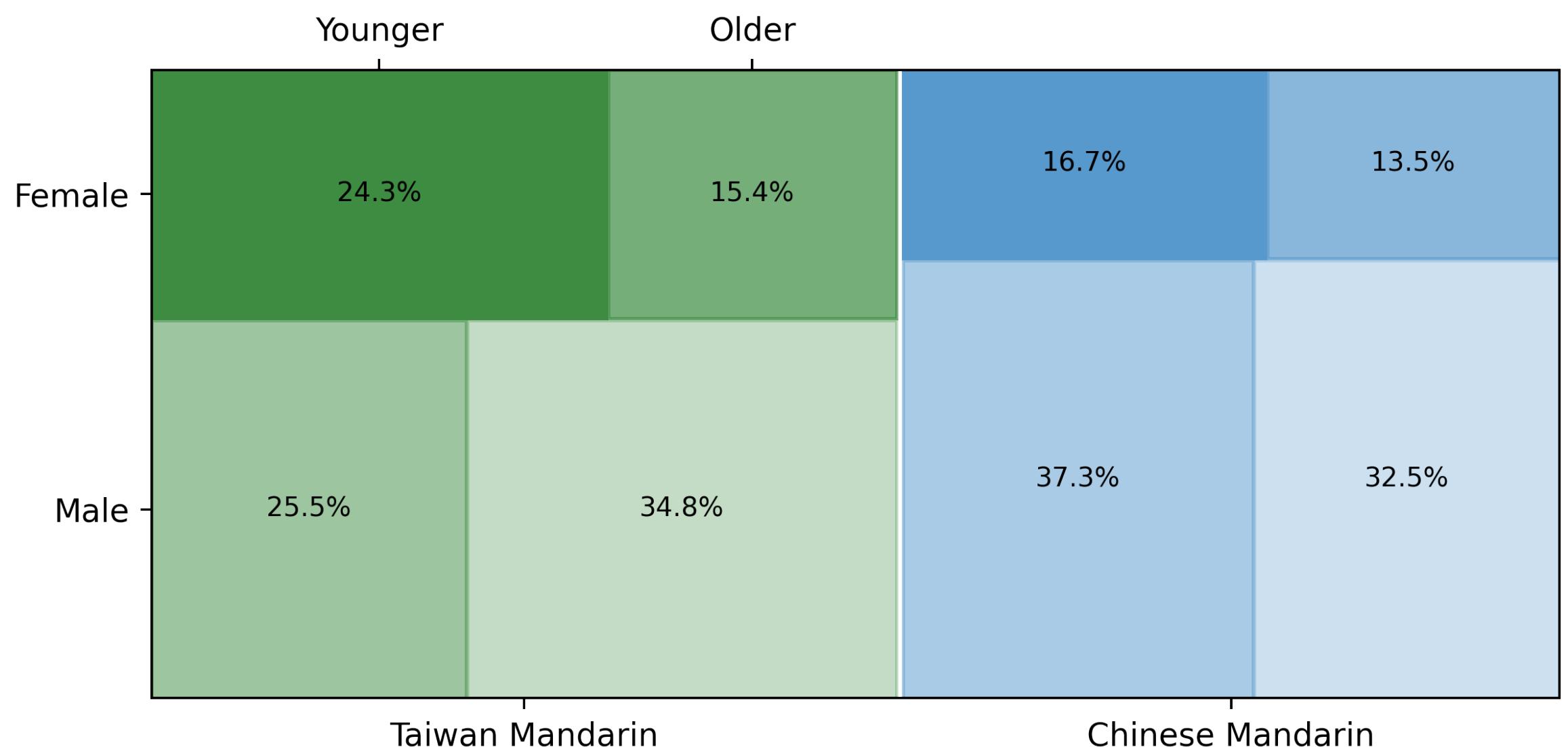
Three pro-Taiwan content creators discussing their views on marriage (范琪斐的美國時間, 2025, Apr. 16).



Methods

Social factor labeling

- Gender & age as control variables were automatically labeled through speech recognition:
 - Gender: Female vs. male
 - Age: 20-39, 40-59
 - A pretrained demographic speech recognition model (Burkhardt et al., 2023) based on Wav2Vec2 (Baevski et al., 2020) was used.
 - To enhance the model's accuracy, the model was further finetuned through active learning with 100 iterations, judged by the first author.





Methods

Statistical analysis

- To account for both levels of classifier specificity (i.e., general vs. specific & continuous PMI), a multivariate Bayesian hierarchical model was fitted:
 - Two linked components:
 - General vs. specific (Bernoulli distribution)
 - Continuous PMI value (Gaussian distribution)
 - Random intercepts for video identity & noun
- The marginal effects of predicted classifier specificity were examined by computing the expected PMI, weighted by each observation's model-predicted probability of using a specific (vs. general) classifier.

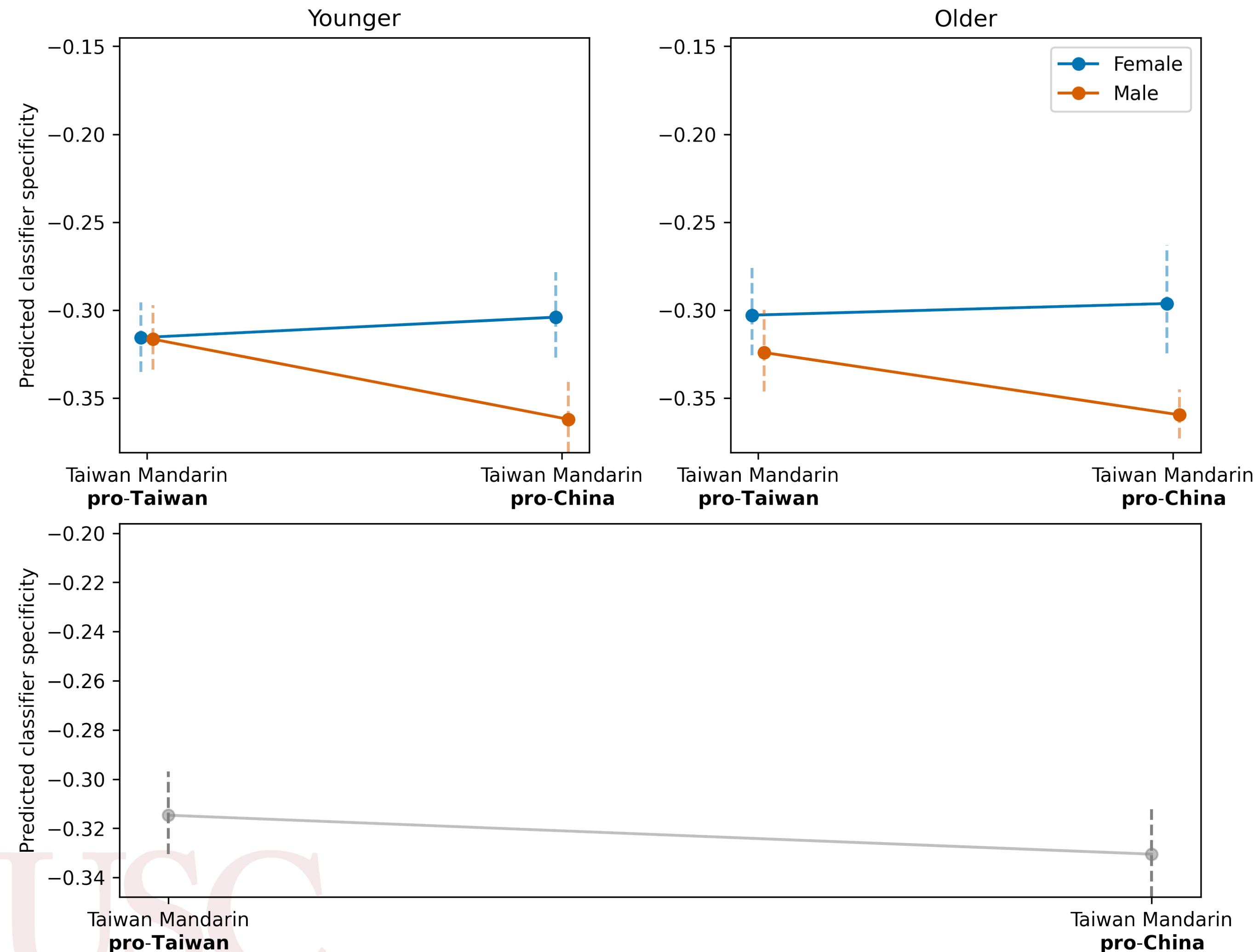
Predicted classifier specificity =

$$PMI_{specific} \times P(specific) + PMI_{general} \times (1 - P(specific))$$



Results

Pro-Taiwan vs. Pro-China



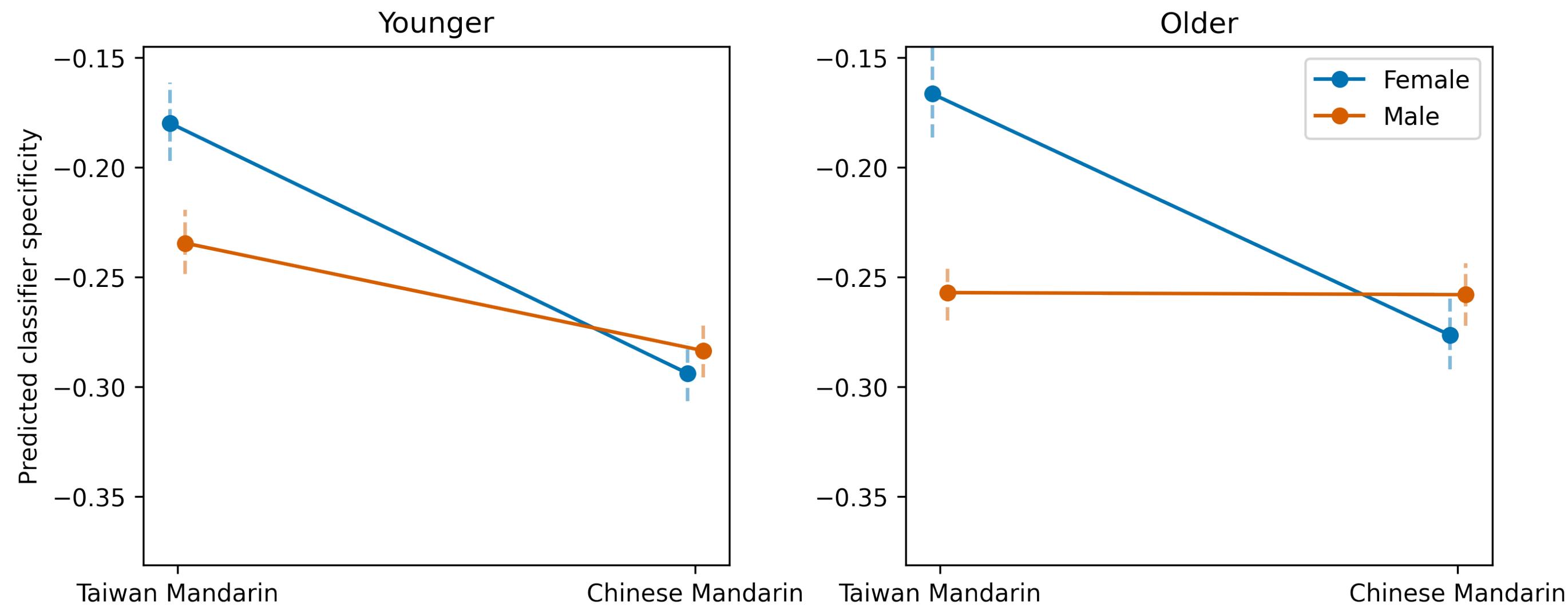
- **Women showed greater classifier specificity** than men.
 - This difference was greater in **pro-China** speakers.
- **Younger** speakers showed **greater classifier specificity**.
- **Pro-Taiwan** Taiwan Mandarin speakers showed **greater classifier specificity** in their production than their **pro-China** counterparts.



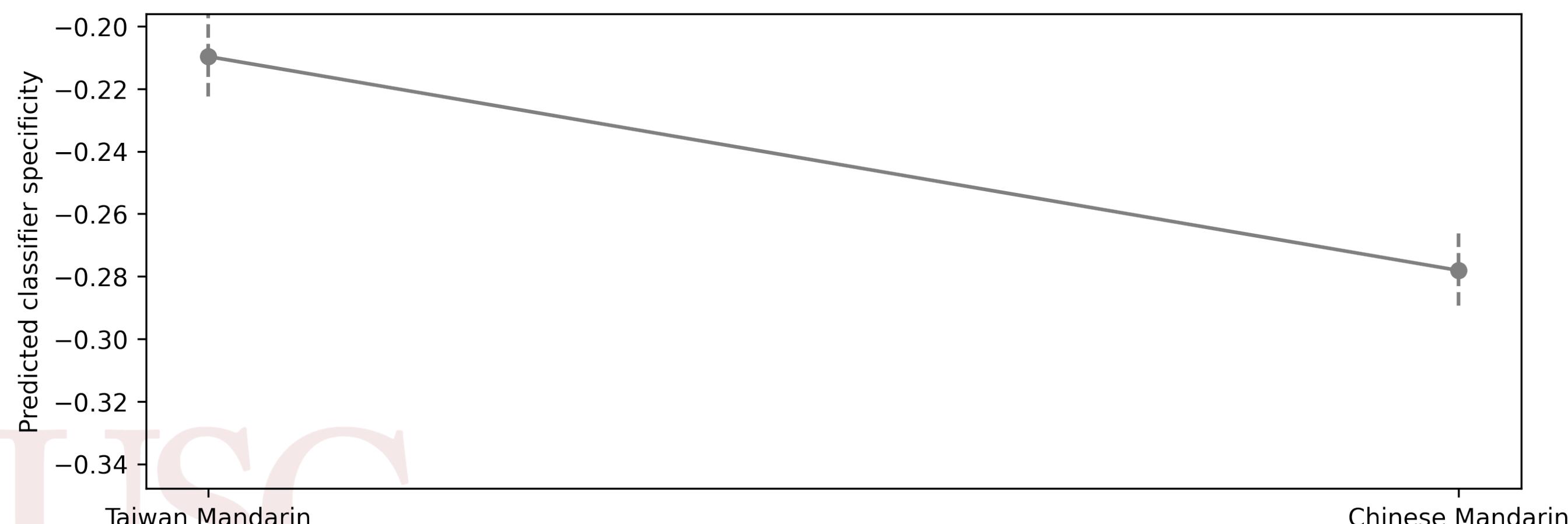


Results

Taiwan Mandarin vs. Chinese Mandarin



- In Chinese Mandarin,
 - Men showed **greater classifier specificity** than women.
 - Older speakers showed **greater classifier specificity**.
- **Taiwan Mandarin** speakers, in general, showed **greater classifier specificity** than **Chinese Mandarin** speakers.

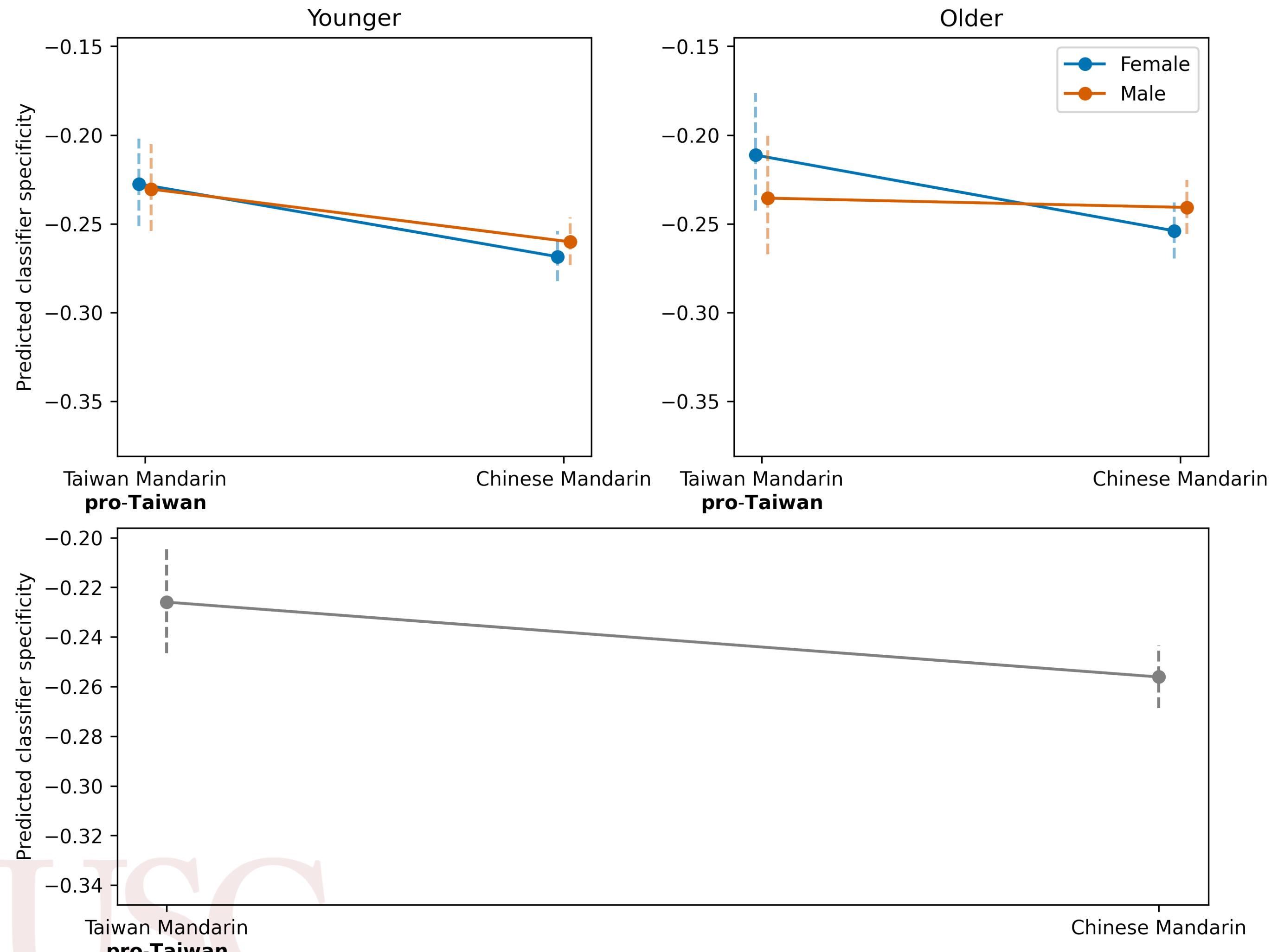


USC



Results

Taiwan Mandarin (pro-Taiwan) vs. Chinese Mandarin



- When compared separately, **both pro-Taiwan & pro-China Taiwan Mandarin speakers showed greater classifier specificity than Chinese Mandarin speakers.**

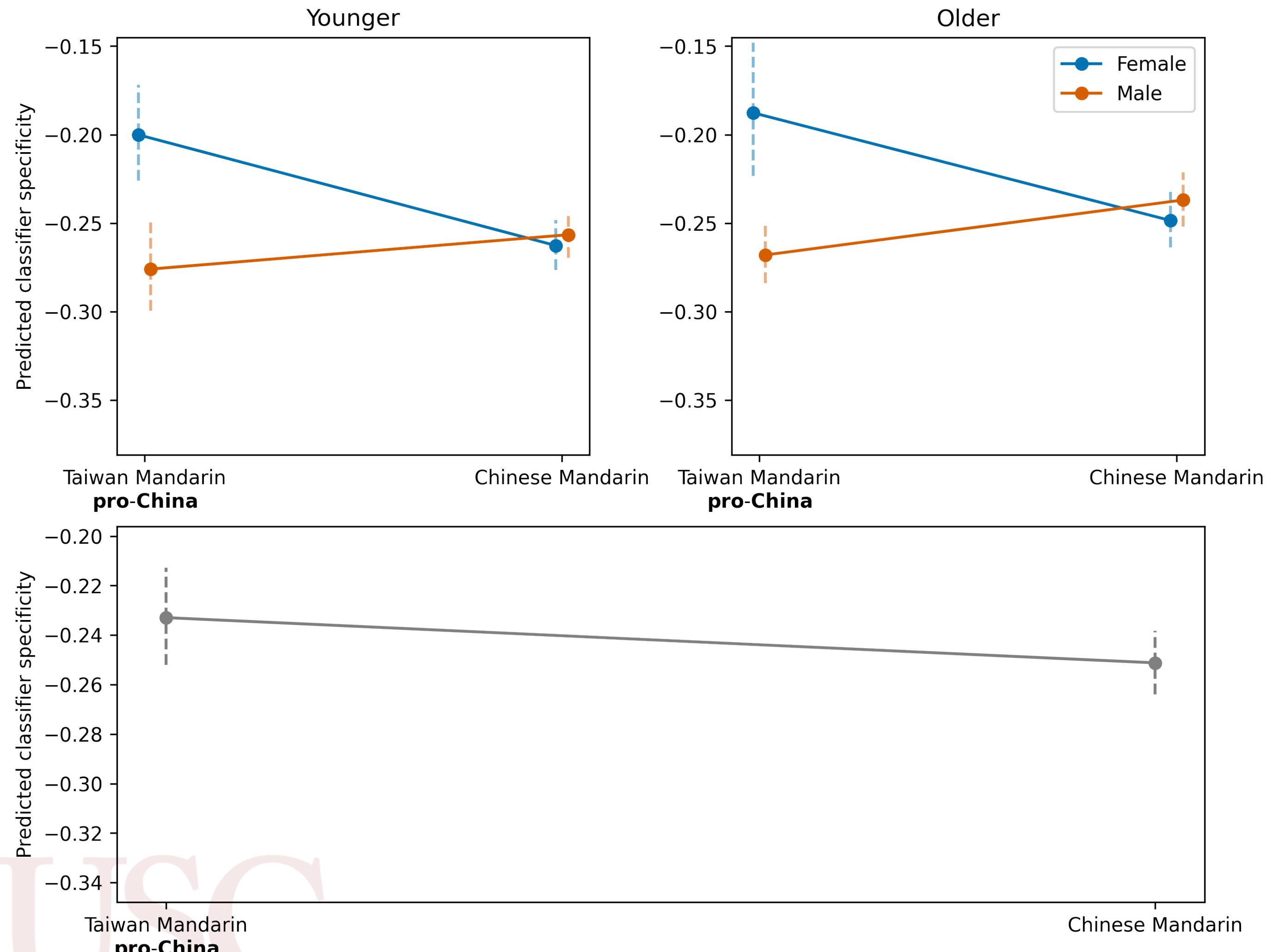


USC



Results

Taiwan Mandarin (pro-China) vs. Chinese Mandarin



- When compared separately, **both pro-Taiwan & pro-China Taiwan Mandarin speakers showed greater classifier specificity than Chinese Mandarin speakers.**



USC



Discussion

The political indexicality of classifier specificity

- Classifier specificity:
 - Taiwan Mandarin (pro-Taiwan) > Taiwan Mandarin (pro-China) > Chinese Mandarin
 - The greater classifier specificity among pro-Taiwan speakers than pro-China speakers suggests the political indexicality of classifier specificity.
 - Similar directionalities were observed b/t pro-Taiwan ~ pro-China & Taiwan Mandarin ~ Chinese Mandarin speakers.
 - Variety differences are likely the target.



Discussion

Female & young speakers as potential change leaders

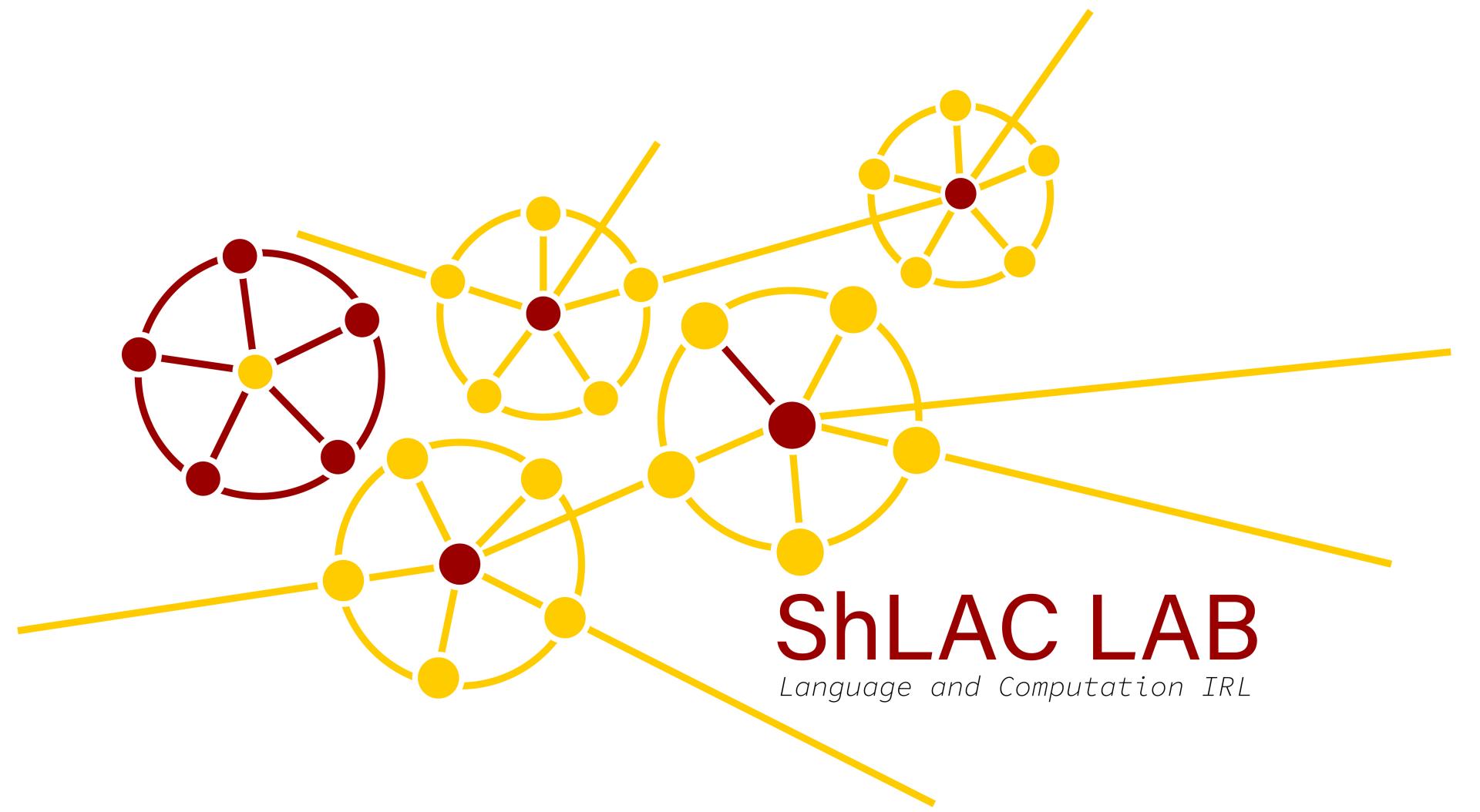
- Within both Taiwan Mandarin & Chinese Mandarin, female & young speakers were aligned (both using either the more or less specific classifiers).
- A unified account could suggest they are change leaders in both languages, w/ opposite directions.



Conclusion

- This study examines the morphosyntactic aspects of political indexicality.
 - Previous studies focused on **lexical and phonetic** variables.
 - Few studies have shown the same politicizability in nuanced **morphosyntactic** variables as classifier specificity.
- We show that the political indexicality of classifier specificity likely exists across the political identity of pro-Taiwan & pro-China Taiwan Mandarin speakers.
- Such indexicality is likely targeting the variety differences b/t Taiwan Mandarin & Chinese Mandarin.

Thank you!



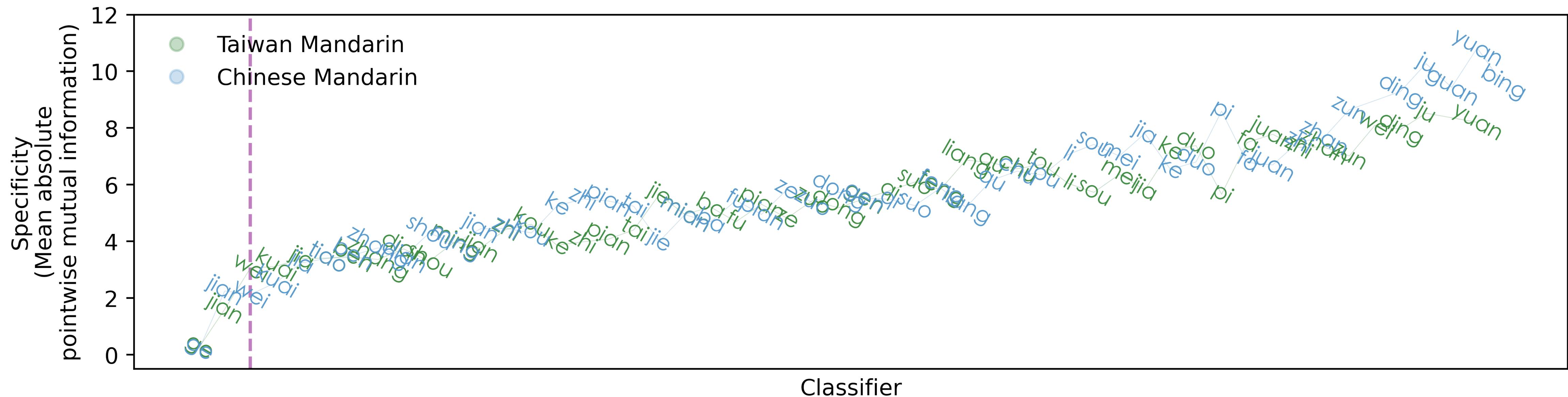
The Taiwan/Beijing
Youtube Corpus



Presentation slides



PMI distribution



USC



Discussion

Gender differences

- Assuming specific classifiers are more formal, the observed greater classifier specificity in female speakers in **Taiwan Mandarin** is in alignment w/ the widely attested gender effects on variation, where female speakers tend to use the more formal variants.
- The weaker gender differences in **pro-Taiwan Taiwan Mandarin** speakers could potentially be due to the generally stronger liberalism among **pro-Taiwan Taiwanese**.
- This account does not explain the reversed gender effects in **Chinese Mandarin**.



Discussion

Age differences

- The observed greater classifier specificity in older speakers in **Chinese Mandarin** is also in alignment w/ the widely attested age effects on variation.
- The reversed age effects (younger speakers showing greater specificity) could be due to the nature of **Taiwan Mandarin**:
 - Her (2009) argues that **Taiwan Mandarin** is a creole influenced by Beijing Mandarin (in the 1940s) & the other Chinese languages (mainly Taiwan Southern Min).
 - The older speakers in this study fall into the first generation who speak this newborn creole. → The regularization of **Taiwan Mandarin** may still be in progress for these speakers.