Assessing gender impact on paralinguistic accommodation in French WhatsApp conversations

Prakhar Gupta, Elisa Pellegrino, Leyla Benkais, Aris Xanthos

Gender and paralinguistic accommodation in WhatsApp

Paralinguistic Cues

- Para-("Besides") + linguistic (language)
- Vocal and (sometimes non-vocal) signal accompanying the linguistic message or speech[1], e.g.,
    - Volume
    - Intensity
    - Pitch
    - Rhythm
    - Facial expression
    - Gestures
    - …

- Allows to emphasize and express emotions

Paralinguistic Cues and CMC

- Hard to use vocal signals in computer-mediated communication (CMC)
  - Use different cues

- Different paralinguistic cues for CMC -
  - Repeated punctuations
  - Letter repetitions
  - Emoticons
  - Emojis
  - Stickers
  - GIFs
  - ....

Paralinguistic Cues and impact of gender

- Gender influences the meaning and usage of para. cues[4][5][6]
  - may lead to accommodative adjustments between dialogue partners
    - resulting in increased or decreased similarity in paralinguistic cues[7]

  - e.g., mutual paralinguistic cue accommodation in conversational settings stronger for mixed gender pairs as compared to same gender pairs[8]

Paralinguistic Cues in CMC and the impact of Gender

- Similar accommodation trends as those observed in F2F conversations
  - Dutch teenage boys and girls show similar paralinguistic style(synchronic) in mixed gender chats[9]

- Does similar accommodation exist in older age pairs?

- Are different para. cues accommodated differently w.r.t. age & gender group?

Our Work

- Examination of dynamic paralinguistic accommodation patterns on a French WhatsApp conversation corpus

- Paralinguistic features studied -
  - Emoji usage (😂, 😃, 🧘🏽‍♂️, 🌍, 🌦️, 🥖, 🚗, 📱, 🎉, ❤️, ✅)
  - Alphabetic character Repetitions ('NOooooo', 'Pierrrrrre','hmmmmm')
  - Punctuation Sequences ('!!', '??','...','?!?')

- Statistical Analysis studying the effect of age and gender on accommodation patterns

Dataset

- Collection of WhatsApp Chats

- Chat members - French speakers residing in Switzerland

- Collected in 2022 by means of a public campaign using a web platform

- Chats anonymized to protect user privacy

Dataset

- Group chats are discarded
- Age groups -
  - 18-34 yo.
  - 35-49 yo.

Nature \ Age	18-34 yo.	35-49 yo.	Total (18-49 yo.)
Same Gender (SG)	261469 (19)	9351 (11)	275778 (33)
Mixed Gender (MG)	40968 (4)	23800 (9)	66153 (15)
Total	302437 (23)	33151 (20)	341931 (48)

Number of chat messages and chats for each age group and gender combination

Methodology

- For a para. cue **X** and user **u** and a given chat, define

$$P(X, u) = \frac{|\text{Messages by } u \text{ with } X|}{|\text{Messages by } u|}$$

- Similarly define

$$P(X, u | X, u') = \frac{|\text{Messages by } u \text{ with } X \text{ in response to a message by } u' \text{ with } X|}{|\text{Messages by } u \text{ in response to a message by } u' \text{ with } X|}$$

where **u'** is the other user

Methodology

- Difference due to accommodation for user *u* and feature *X* is defined as (Danescu-Niculescu-Mizil et al., 2011)

$$diff(X, u) = P(X, u | X, u') - P(X, u)$$

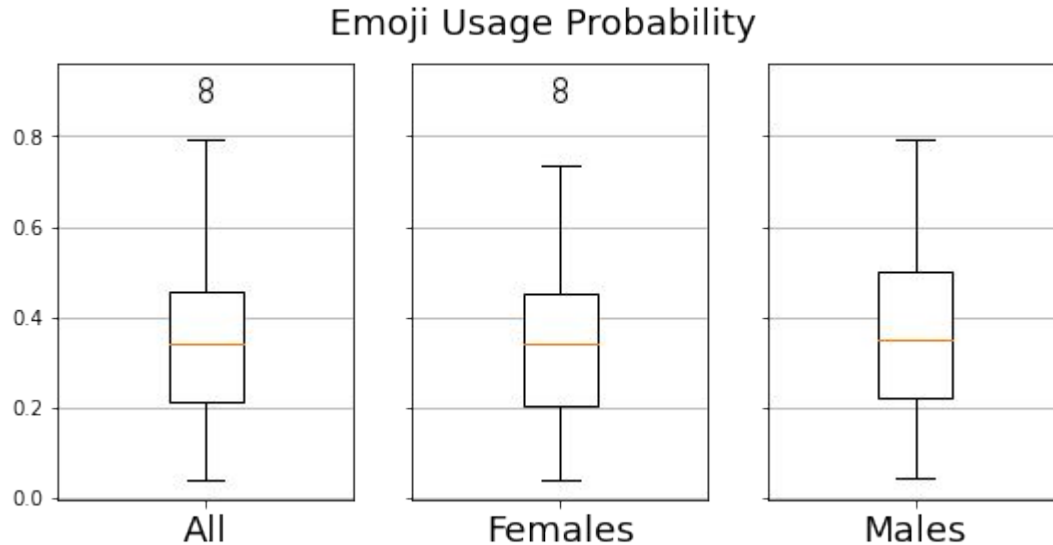- Amplification due to accommodation for user *u* and feature *X* is defined as

$$ampl(X, u) = \frac{P(X, u | X, u')}{P(X, u)}$$

Results - Basic Statistics

Paralinguistic Cue	Avg. Probability over chats
Emojis	0.346
Character Repetition	0.069
Punctuation Sequences	0.139

Results - Emojis

- Emoji usage distribution similar w.r.t. gender


Emoji Usage Probability

Results - Emojis

- Difference due to accommodation significantly higher than 0 (p < 0.05) for all age-group and gender combinations

- Amplification due to accommodation significantly higher than 1 (p < 0.05) for all age-group and gender combinations

Results - Emojis

- Amplification due to accommodation significantly higher in same gender than mixed gender chats (p = 0.038)

Gender and paralinguistic accommodation in WhatsApp
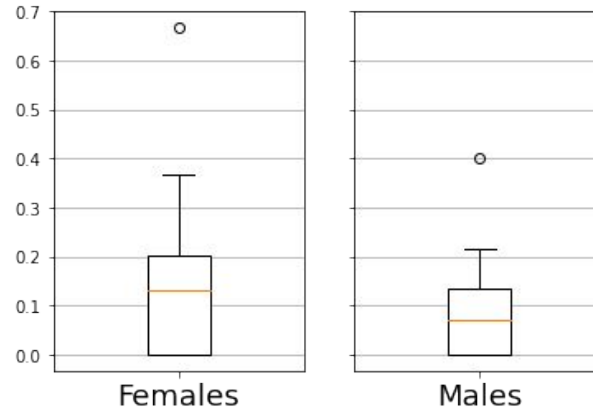
Results - Character Repetitions

- Character repetition distribution is dissimilar w.r.t. gender
  - Females show more character repetition usage than males
  - But not statistically significant (p = 0.101)

Character Repetition Occurrence Probability
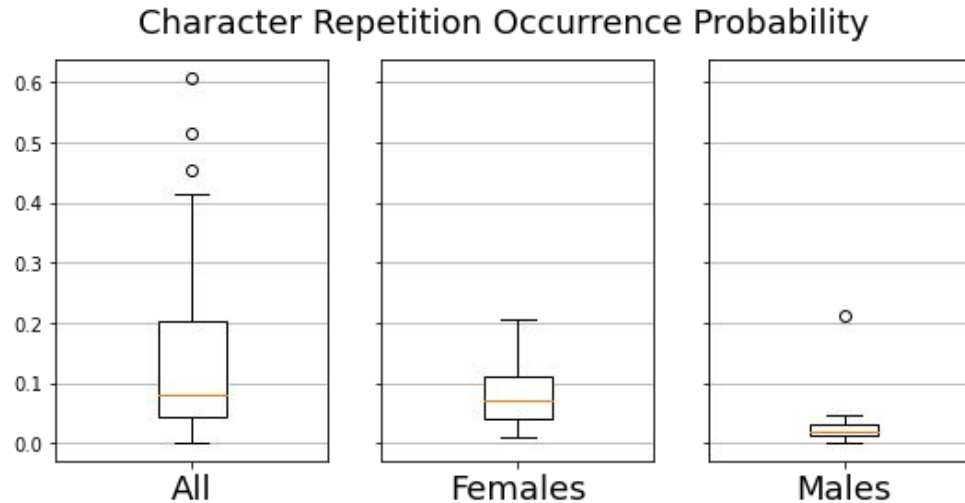
Results - Character Repetitions

- Distribution for character repetition as a response to character repetition is comparatively less dissimilar w.r.t. gender
  - females > males
  - statistically not significant (p = 0.559)

Response with char. repetition to a message with char. repetition probabilities
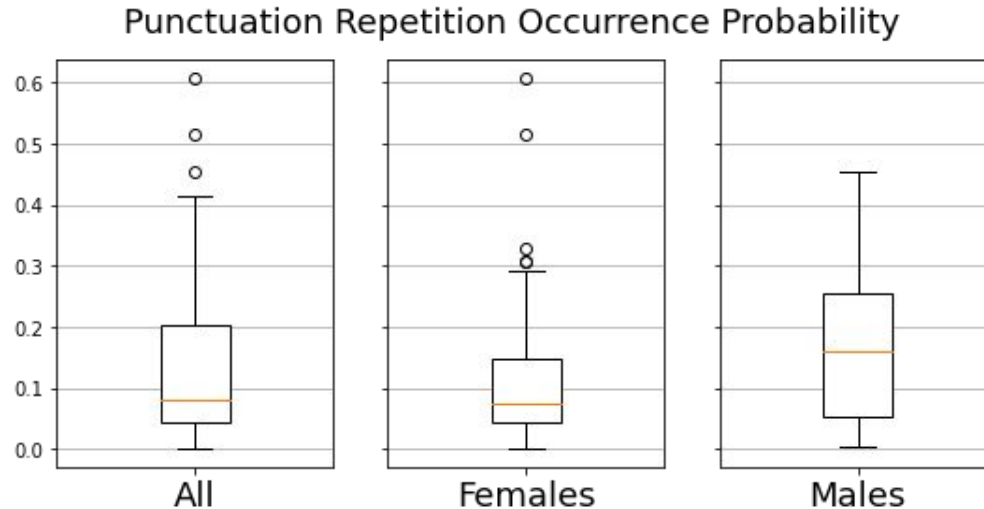
Results - Character Repetitions

- Difference due to accommodation significantly different from 0 & amplification significantly different from 1 only for
  - the complete dataset and,
  - same gender chats

Results - Character Repetitions

- Amplification and difference due to accommodation significantly higher in same gender than mixed gender chats
  - p = 0.016 for difference
  - p = 0.048 for amplification

Gender and paralinguistic accommodation in WhatsApp
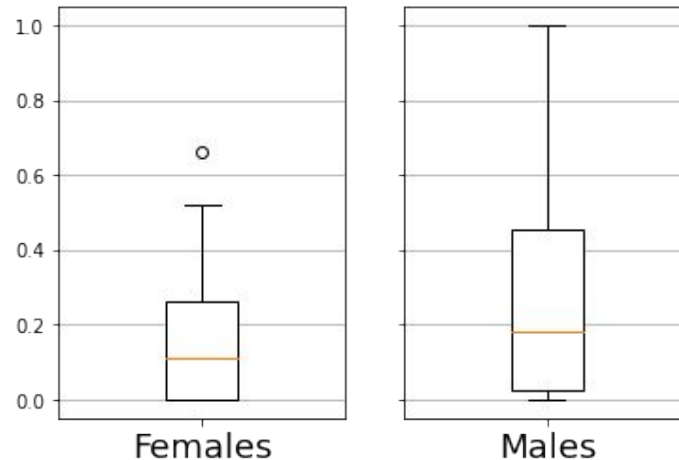
Results - Punctuation Sequences

- Punctuation sequences distribution is dissimilar w.r.t. gender
  - Males show more punctuation sequence usage than females
  - But not statistically significant ($p = 0.145$)

Punctuation Repetition Occurrence Probability

Results - Punctuation Sequences

- Punctuation sequences accommodation distribution is dissimilar w.r.t. gender
  - Males show more punctuation sequence accomodation than females
  - But not statistically significant ($p = 0.059$)

Response with punct. seq. to a message with punct. seq. probabilities

Results - Punctuation Sequences

- Difference due to accomodation significantly higher than 0 &
  Amplification due to accomodation significantly higher than 1 only for
  - the complete dataset and,
  - mixed gender chats

Results - Punctuation Sequences

- Difference due to accommodation significantly higher when comparing mixed gender to same gender chats ($p = 0.021$)

Summary

- Users show dynamic accommodation of paralinguistic cues, i.e.,
  - use all three paralinguistic cues significantly more frequently in response to messages containing them

- Effect of gender composition
  - Mixed gender pairs show more dynamic accommodation of the punctuation repetition
  - Same gender pairs show more dynamic accommodation of the character repetition and emojis

- No significant effect of age group found within our data

Limitations

- Small Dataset

- Geographical and linguistic restrictions
  - Most of the speakers originate in Swiss Romandy
  - Only French is studied

- Only chats with cis-gender speakers were considered

- Effect of relationship b/w the speakers not taken into account

References

1. Trager, George L. "Paralanguage: A first approximation." Stud. Linguist. 13 (1958): 1-12.
2. Carey, John. "Paralanguage in computer mediated communication." 18th Annual Meeting of the Association for Computational Linguistics. 1980.
3. Oberlander, Jon, and Alastair J. Gill. "Language with character: A stratified corpus comparison of individual differences in e-mail communication." Discourse processes 42.3 (2006): 239-270.
4. Chrisler, Joan C., and Donald R. McCreary. Handbook of gender research in psychology. Vol. 1. New York: Springer, 2010.
5. Yu, Alan CL. "Perceptual cue weighting is influenced by the listener's gender and subjective evaluations of the speaker: the case of English stop voicing." Frontiers in Psychology 13 (2022): 840291.
6. Dijksterhuis, A., & Bargh, J. A. (2001). The perception–behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), Advances in experimental social psychology, Vol. 33 (pp. 1–40). Academic Press.
7. Pardo, J. S., Pellegrino, E., Dellwo, V., Möbius, B. (2022). Special issue: Vocal accommodation in speech communication. Journal of Phonetics, 95, 101196, ISSN 0095-4470.
8. Levitan, R., Gravano, A, Willson, L.,Benus, S., Hirschberg, J. and Nenkova, A. (2012). Acoustic-Prosodic Entrainment and Social Behavior. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 11–19, Montreal, Canada, June 3-8, 2012.
9. Hilte, L., Vandekerckhove, R. & Daelemans, W. (2022). Linguistic Accommodation in Teenagers' Social Media Writing: Convergence Patterns in Mixed-gender Conversations, Journal of Quantitative Linguistics, 29(2): 241–268.