From Thoughts to Words: Exploring Misarticulation in Texts

Sophie Wang

University of Chicago

17 May 2022

Misarticulation

- A newly recognized psychological mechanism
- High frequency in daily life
- **Definition:**

  The subjective sense of not having fully expressed or communicated what one has in mind.

Background

- Existing literature:

    underlying process of information transmission in communications

- Misarticulation:

    psychological experience of the communicators themselves

Background

- Humans are self-interpreting animals

- Function of words

- Perspectives of Miscommunication
  - in association with interpersonal relations
  - problems in the physical transmission of messages
  - features of the social context of interaction
  - **psychological perspective** of the communicator themselves (new)

Research Question

- Whether misarticulation is predictable through texts

- Why exploring this?
  - Contribute to the knowledge of (mis)communication
  - An interdisciplinary study about a novel concept

Methods

- Online experiment designed using Qualtrics

Process:
1. Consent form previously approved by the IRB
2. A set of demographic questions
3. Randomly assigned to one of the three situations
4. Think and write responses around 140 words
5. Rate for three questions --> misarticulation index

Methods

- 524 participants hired from Amazon Turk
- Text pre-processing: Python NLTK library
- Text vectorization: word embedding - - GloVe
- Model training: Python Scikit-learn library

Methods

- **Analysis 1**

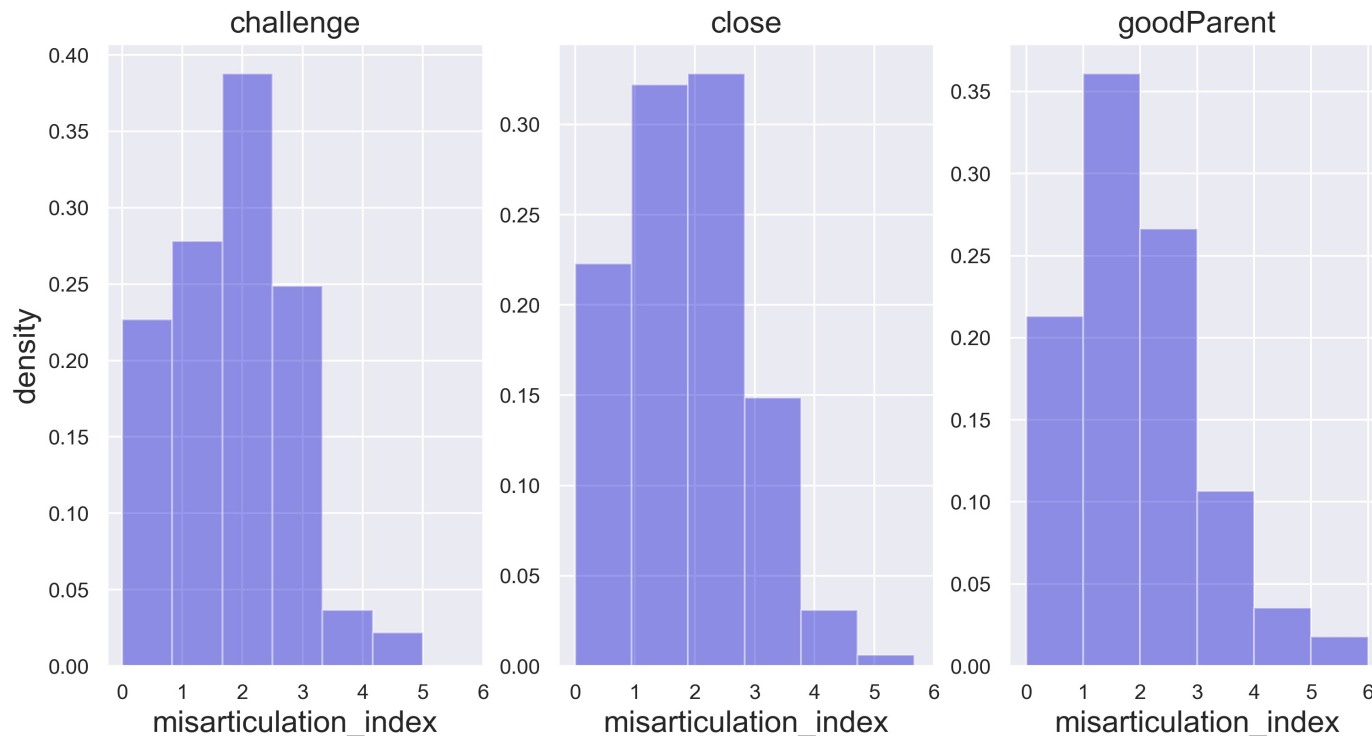   Prediction model: Support Vector Machine Regression

- **Analysis 2**

    Classification model: Logistic Regression

    + 5- fold cross validation

Analysis 1 – Data

- **Input X**: textual responses from each participants (N = 504)
- **Label y**: misarticulation index
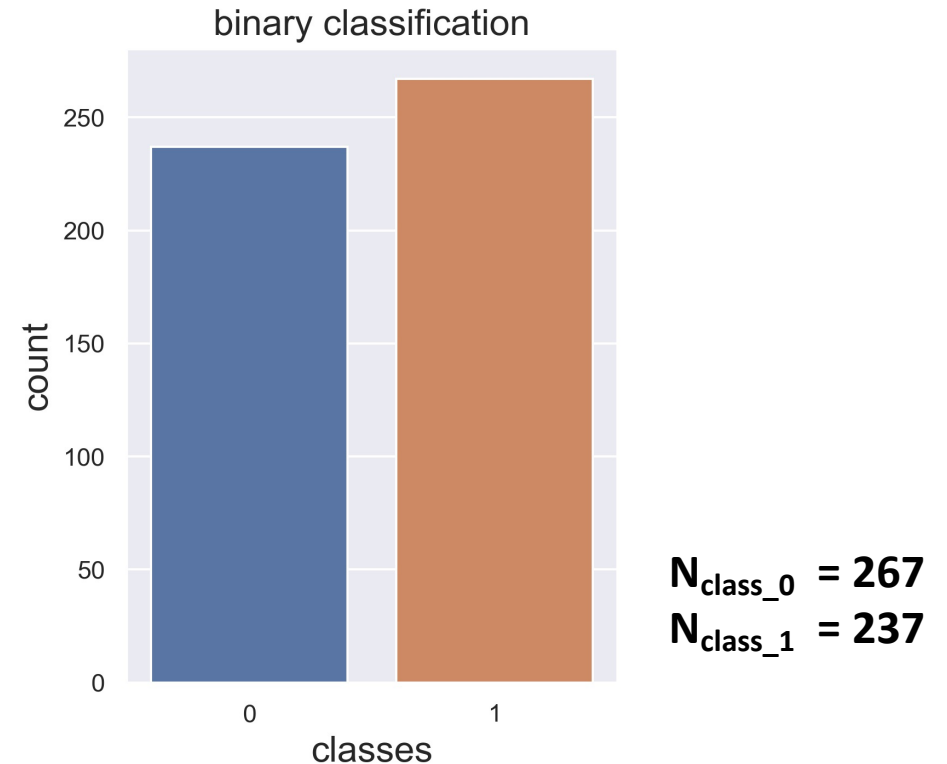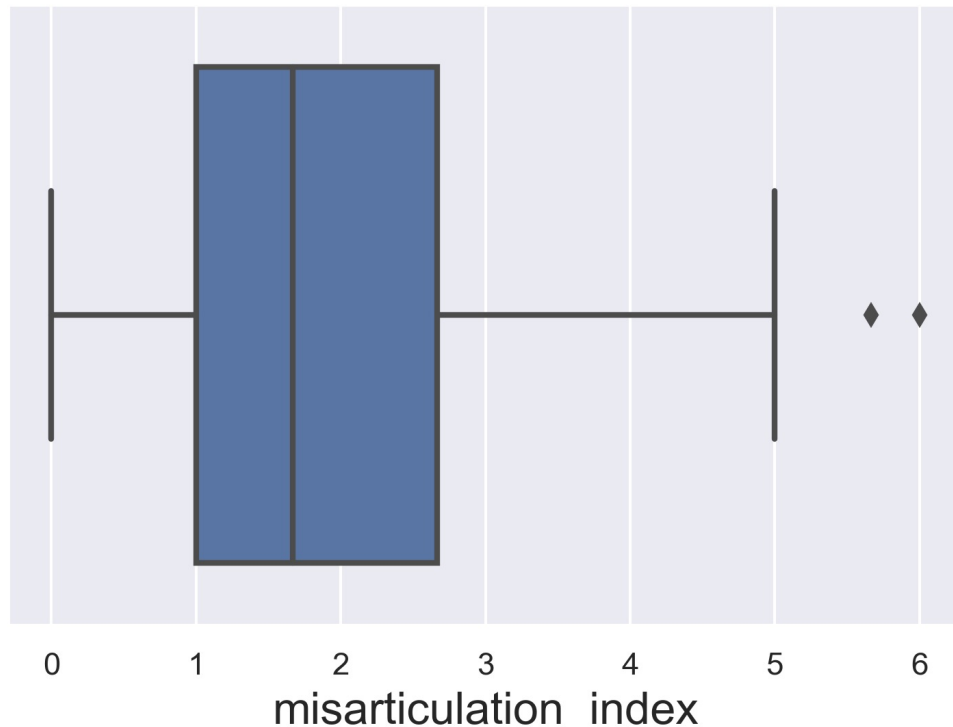
Analysis 1 – Findings

- Benchmark: random shuffling prior to model training

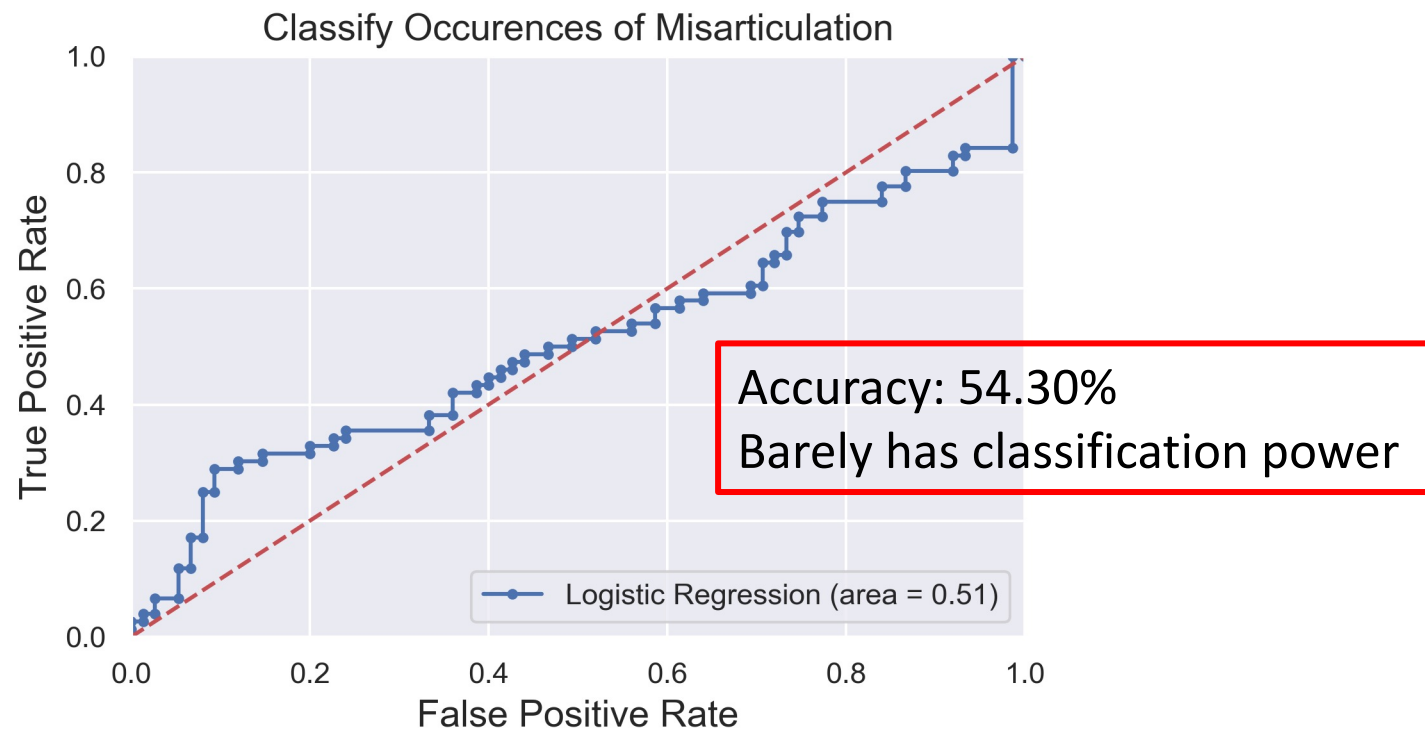|	SVR model	random vector shuffle for X	random true value shuffle for y
Mean Absolute Error (MAE)	0.88	0.88	0.92
Mean Squared Error (MSE)	1.26	1.21	1.20

Does not outperform the benchmark

Analysis 2 – Data

- **Input X**: textual responses from each participants (N = 504)
- **Label y**: binary classes (class 0 & class 1)

$N_{class\_0} = 267$
$N_{class\_1} = 237$

Analysis 2 – Findings

- Benchmark: split on the mean value (94% accuracy?)
  - N_{class_0} = 453, N_{class_1} = 51

Classify Occurences of Misarticulation

Accuracy: 54.30%
Barely has classification power

Logistic Regression (area = 0.51)

Discussion

- Misarticulation can hardly be predicted through texts…
  - in this current study – an exploration

- There are still possibilities for future research on misarticulation

Limitations

- Quality of generated data
  - Perhaps it just naturally doesn't have enough features to signal misarticulation

- Feature representation
  - Pre-trained word embedding model may accidentally filter out important features
  - More sophisticated method may help

References

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? Cognition, 96(3), 263–284. https://doi.org/10.1016/j.cognition.2004.09.002

Jackson, J. C., Watts, J., List, J.-M., Puryear, C., Drabble, R., & Lindquist, K. A. (2021). From Text to Thought: How Analyzing Language Can Advance Psychological Science. Perspectives on Psychological Science. https://doi.org/10.1177/17456916211004899

Levelt, W. J. M. (1999). Models of word production. Trends in Cognitive Sciences, 3(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4

Smith, K.E. (2010). "Chapter One. The Meaning of Meaning". In *Meaning, Subjectivity, Society*. Leiden, The Netherlands: Brill. doi: https://doi-org.proxy.uchicago.edu/10.1163/ej.9789004181724.i-266.7

Thank you for listening!