Employing Machine Learning and Language Models to Differentiate Language Patterns in Mandarin-Speaking Preschoolers with Autism Spectrum Disorder

Yue Chen[1], Ziyin Mai[2], Yige Chen[2]

[1] Department of Linguistics, University of Southern California

[2] Department of Linguistics and Modern Languages, Chinese University of Hong Kong

SCAN ME!

Multilingual input & child language development
Research projects funded by Research Grants Council, HKSAR
2021-2025, GRF #14615820, ECS #21604522

Background

- Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by significant difficulties in communication and language development [Baird and Norbury, 2016].
- Mandarin-speaking preschoolers with ASD may struggle with lexical and grammatical skills, often producing shorter Mean Length of Utterance in words (MLUw) and reduced lexical diversity (Type) compared to their typically developing (TD) peers [Su et al., 2018, Zhou et al., 2015].
- Despite this, research on identifying ASD in preschoolers using machine learning (ML) and language models (LM) based on naturalistic speech data remains limited.
- This pilot study explores whether ML and LM can **differentiate between ASD and TD utterances**, potentially providing an **ecologically valid, cost-efficient screening tool** for early ASD intervention.

Research Question

- How effectively can machine learning models and large language models (LLMs) differentiate language patterns in child speech between typically developing (TD) Mandarin-speaking preschoolers and those with Autism Spectrum Disorder (ASD)?

Dataset

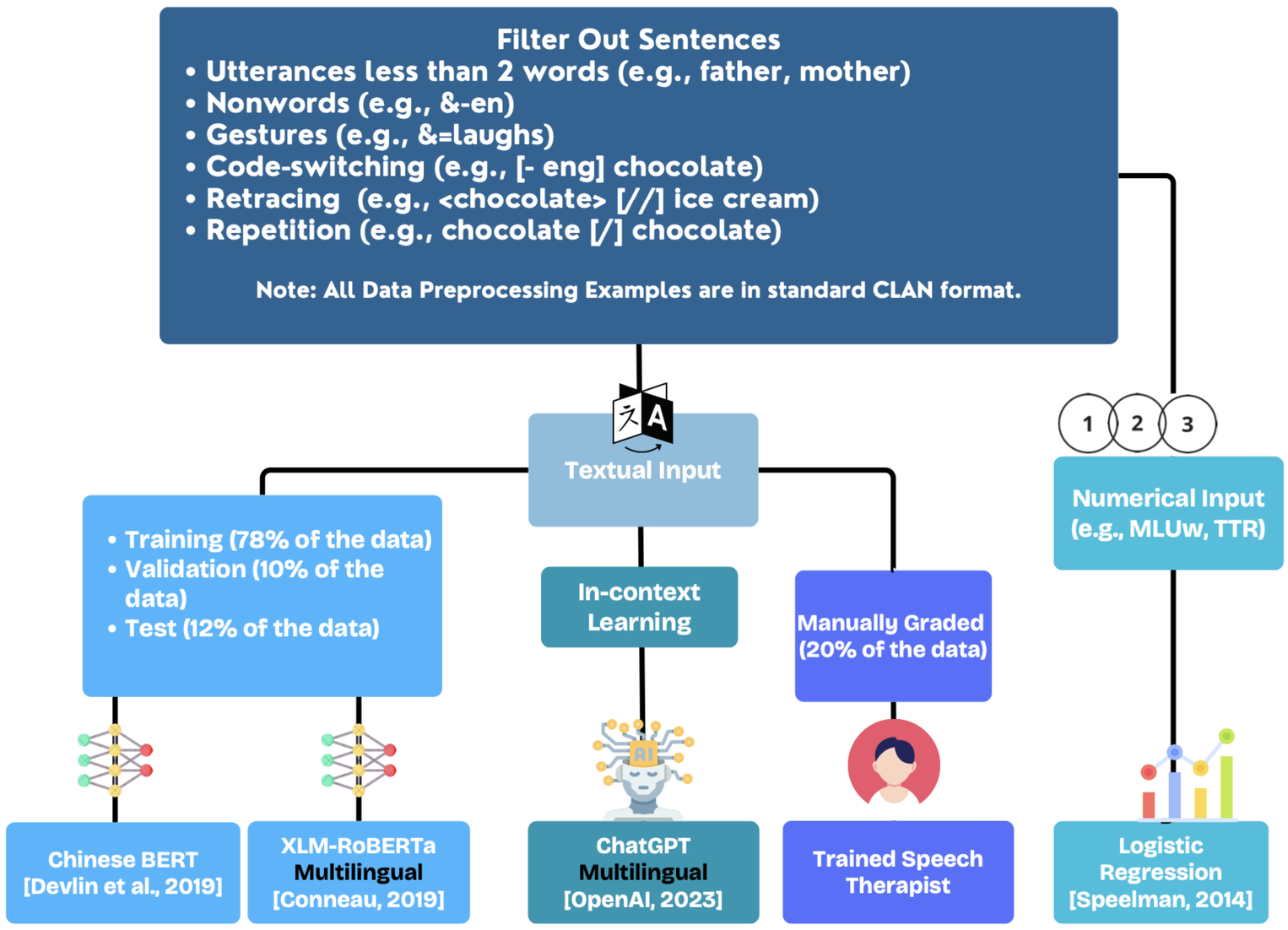
Adult-Child Interaction

ASD	TD
Mandarin Shanghai Corpus [Zhou, 2006]	**Beijing Child Mandarin Corpus** [Mai, Prep]
Longitudinal	Cross-sectional
AgeRange = 4;0 - 6;0	**AgeRange** = 3;0 - 6;0
N = 7	**N** = 43
N_Utterance = 2975	**N_Utterance** = 3016
MLUw = 3.95	**MLUw** = 6.21
Type = 72.29	**Type** = 155.39

Acknowledgments

We are deeply grateful to our colleagues and collaborators: Jingyao Liu, Xuening Zhang, Yuqing Liang, and Jiaqi Nie. We are also grateful to Dr. Elsi Kaiser and Dr. Zuzanna Fuchs from USC, for their invaluable suggestions and comments. The Research grants awarded to Ziyin Mai: "Input and experience in early trilingual development", RGC/GRF, 2021-2024; "Input and caretaker proficiency in early bilingual development: mothers, helpers and toddlers ", RGC/ECS, 2023-2025.

Data Preprocessing



Results

- Our results show that the utterances produced by children with ASD typically exhibit a shorter Mean Length of Utterance in words (MLUw) and reduced lexical diversity, aligning with previous research findings [Rice et al., 2010, Sandbank and Yoder, 2016].
- A significant difference in **MLUw ($p < 0.05$)** was observed between children with and without ASD, indicating that **higher MLUw** is associated with a **lower likelihood of ASD**.
- All models demonstrated high accuracy, precision, and strong reliability in classifying utterances produced by children with ASD, with the exception of **ChatGPT** (see Table 2).
- The speech therapist's ratings showed lower accuracy, likely due to the use of different criteria, such as considering a stereotyped and idiosyncratic word or phrase usage across utterances, **behavioral gesture data**, and disfluency, which were not included in the judged utterances.

Table 2: Performance Comparison

Models	Description	Data Type	Basic Unit	Test Set (Full, 12%)		Test Set (Partial, 2.4%)	
				Accuracy	Macro F1	Accuracy	Macro F1
BERT	Encoder-only Language Model	Textual	Utterance	0.89	0.88	0.93	0.93
XLM-RoBERTa	Encoder-only Language Model	Textual	Utterance	0.88	0.87	0.92	0.92
ChatGPT	Decoder-only Language Model	Textual	Utterance	0.49	0.46	0.48	0.47
Speech Therapist	Human Rater	Textual	Utterance	NA	NA	0.62	0.61
Logistic Regression	Statistical Machine Learning	Numerical	Transcript	0.90	0.88	NA	NA

Conclusion

- This study lays the groundwork for **developing practical, non-invasive early detection tools** for Mandarin-speaking children in clinical settings.
- The performance gap between the models and human raters is likely due to differences in evaluative criteria. Tools like ChatGPT, similar to human raters, may incorporate broader contextual and behavioral insights when evaluating ASD-related utterances.