I'm Aida Tarighat, a researcher in computational psycholinguistics.

I investigate natural language disfluencies and nonliteral meaning in written language using behavioral experiments and language modeling.
My interdisciplinary work aims to integrate psycholinguistic insights into AI systems for mental health support and user-centered interaction design.

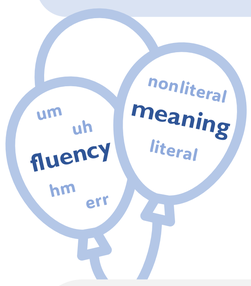
My supervisors are Martin Corley and Patrick Sturt from PPLS.

My academic training spans language and literature, general psychology, and natural language processing.

I'm actively exploring my next research role. Let's connect. Happy to chat about collaborations or research ideas!

This is most of my PhD in an A0-sized nutshell.

Written disfluencies can be *um* slow to read.



- Are they difficult for humans to interpret?
- Can they help signal nonliteral meanings like sarcasm?
- Do language models process them like humans do?

Written disfluencies are rarely experimentally studied. Yet people do use them (confirmed in a previous study).

While humans can interpret these cues, language models often struggle due to training on filtered data. We aim to measure and improve their ability to comprehend and generate nuanced, emotionally informed, and contextually aware dialogue.

why written disfluency?

In speech, disfluencies like *um* or *uh* may be unintentional or strategically used to convey a meaning. But in writing, they're always intentional and potentially meaningful. We build on findings from spoken language and theories like the Graded Salience Hypothesis to explore whether written disfluency might aid comprehension, particularly in accessing nonliteral meanings like sarcasm.

takeaway

human reading:

Written disfluent statements containing *um* were slower to read. Nonliteral meaning also slowed readers down. We didn't find an interaction between fluency and meaning.

human predictions:

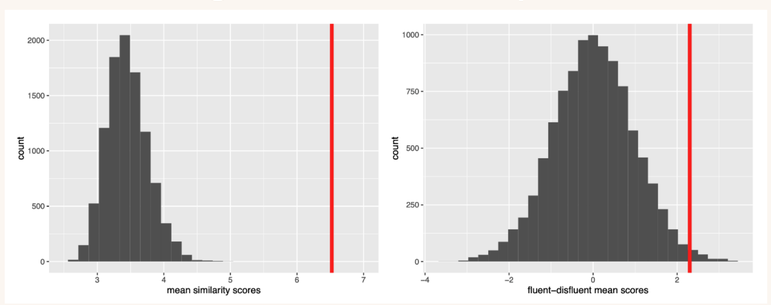
Disfluency slightly affected participants' choice of literal or nonliteral word. Some participants reported tendency to produce sarcastic words when *um* was present.

human evaluation:

Although whatever statement given, disfluency rendered it more sarcastic, already sarcastic statements were rated as less sarcastic in the presence of disfluency. Higher sarcasm scores were rated as more offensive. Fluent statements were perceived as slightly more offensive. Disfluency didn't strongly modulate the effect of sarcasm on offensiveness ratings.

LM's human-like predictions:

BERTweet's continuations were better matches to human continuations following fluent items compared to disfluent items.



Left: red vertical line indicating the mean similarity score of 6.52. Right: red vertical line indicating the observed fluent-disfluent difference in similarity score of 2.30.

methods and measures

Behavioral and language modeling experiments to examine the effect of meaning, fluency, and their interaction:

reading self-paced (99 participants) eye-tracking (59 participants) by assigning target and spillover regions and using reading time and eye gaze in milliseconds `[log-transformed time, linear mixed-effects modeling]`

cloze completion test (160 participants) – asking participants to guess the missing target word to obtain the most popular completion to later compare to our language modeling task `[mean]`

statement evaluation task (101 participants) – asking participants to evaluate the statements on 5-point Likert scales for sarcasm, offensiveness, language formality, and associated emotions [Ekman's 6 basic emotions] `[Bayesian regression modeling using cumulative logit link]`

masked language modeling on BERTweet to obtain the top 10 model predictions for the excised tokens denoting literal and nonliteral meanings in fluent and disfluent versions `[LSA cosine similarities by pairwise comparisons using word2vec (top cloze completion x confidence rating of top model prediction x word2vec similarity score)][permute LSA and BERTweet scores 10k times and recalculate mean similarity]`

causal language modeling on LLMs to obtain model predictions for the excised tokens denoting literal and nonliteral meanings in fluent and disfluent versions – also includes different placements and spellings of *um* `[study in progress, à la MLM task]`

materials in 4 versions according to meaning and fluency
used in both behavioral and language modeling tasks

literal fluent	Sitting through an hour of sermon would make most children **feral** on any day. You can ask them.
literal disfluent	Sitting through an hour of sermon would make most children, um, **feral** on any day. You can ask them.
nonliteral fluent	Sitting through an hour of sermon would make most children **merry** on any day. You can ask them.
nonliteral disfluent	Sitting through an hour of sermon would make most children, um, **merry** on any day. You can ask them.

for context

I became interested in (intentional) disfluency while interviewing psychotrauma survivors for my psychology dissertation. I wanted to explore the potential meaning behind these interruptions. In my research, I'm not aiming to make language models disfluent or sarcastic, but to improve their understanding of psycholinguistic features like written disfluencies and nonliteral meanings.