Umm, I have a strong urge to reply “wait, didn’t you date (name) that year?” to anyone that caption “together since (year)... Anyway, happy bitter day to singles out there.

fillers (a.k.a. discourse markers and filled pauses)

um

hmm

uh

er

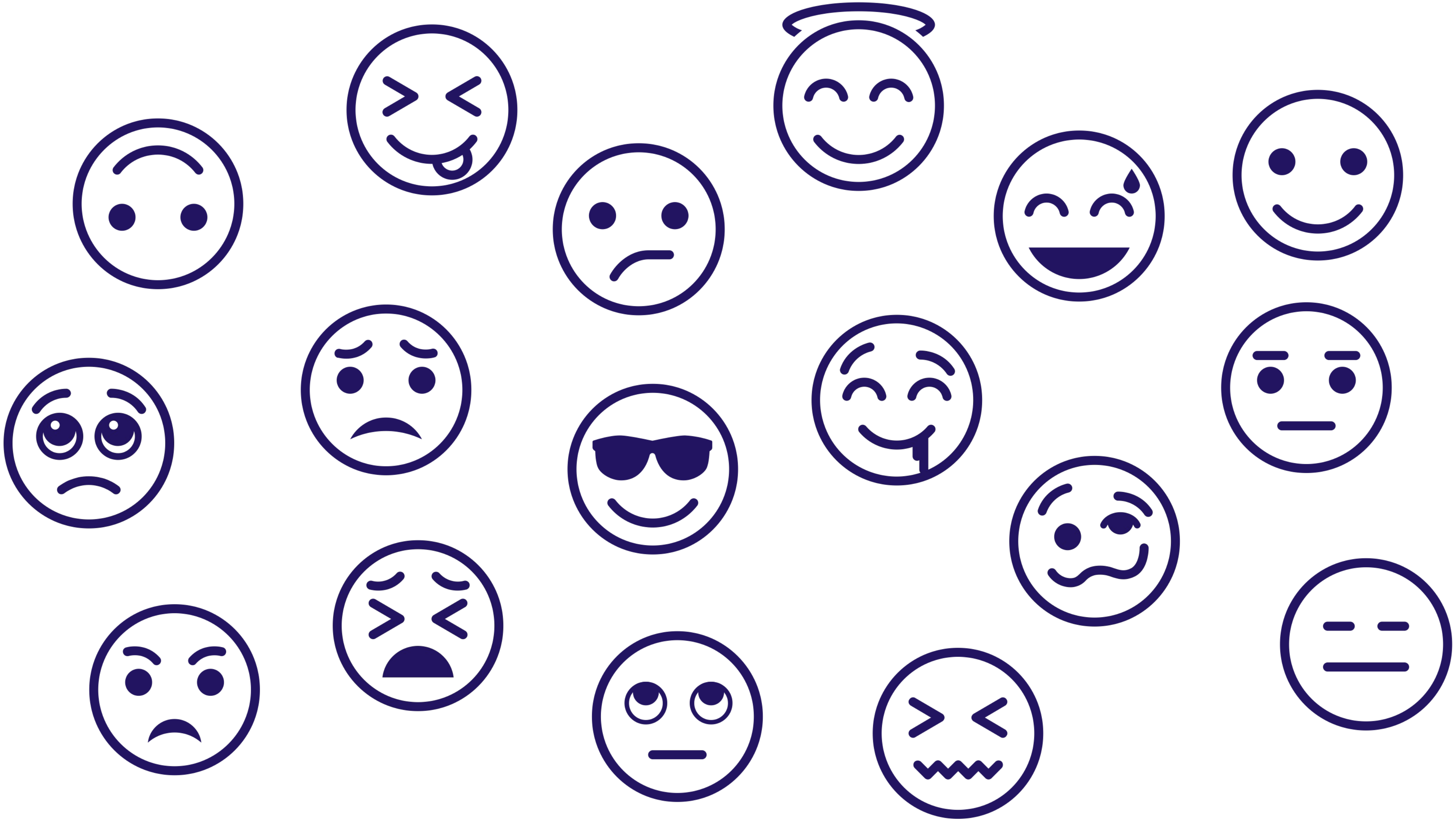
erm

Perspectives on Language Model and Human Handling of Written Disfluency and Nonliteral Meaning

Aida Tarighat

with **Patrick Sturt** & **Martin Corley**

NLP
UKRI CENTRE FOR DOCTORAL TRAINING



When written, disfluencies are intentional.

I coughed up a great amount of blood just now- umm....

I coughed up a great amount of blood just now- umm....

~~prosodic cues~~

~~gestures~~

~~facial expressions~~

Interpreting written fillers is not straightforward.

not using or interpreting words in their typical, literal, or most basic meanings

not using or interpreting words in their typical, literal, or most basic meanings

sarcasm

Written disfluencies are not experimentally studied.

spoken disfluencies can affect interpretation

Fillers like *um* and *uh* speed up the processing of the word which follows them.

Fillers help with the integration of unexpected words into their discourse.

Fillers bias expectations toward new rather than given information.

spoken disfluencies can affect interpretation

Fillers like *um* and *uh* speed up the processing of the word which follows them.

Fillers help with the integration of unexpected words into their discourse.

Fillers bias expectations toward new rather than given information.

fillers could potentially signal nonliteral meaning

Spoken fillers influence listeners’ pragmatic interpretations, guiding them toward particular meanings.

spoken disfluencies can affect interpretation

Fillers like *um* and *uh* speed up the processing of the word which follows them.

Fillers help with the integration of unexpected words into their discourse.

Fillers bias expectations toward new rather than given information.

fillers could potentially signal nonliteral meaning

Spoken fillers influence listeners’ pragmatic interpretations, guiding them toward particular meanings.

fillers are being written online

Graded Salience Hypothesis (Giora, 2003; Giora and Fein, 1999)

Humans have difficulty understanding nonliteral meaning because salient (default) meanings have cognitive priority in language comprehension, and accessing an alternative (such as an ironic or sarcastic interpretation) is cognitively effortful.

We hypothesize

The use of *um* in a sarcastic context (in speaking or in writing) signals an interruption of the salient context, making it easier for listeners or readers to access the intended, nonliteral, meaning.

We want to know

how well LMs could handle written disfluencies,

We want to know

how well LMs could handle written disfluencies,

whether written disfluencies could signal nonliteral meaning,

We want to know

how well LMs could handle written disfluencies,

whether written disfluencies could signal nonliteral meaning,

whether they could influence the ways in which readers interpret what they are reading.

1 Compare LM and human treatment of written disfluency and predicting nonliteral meanings

1. Compare LM and human treatment of written disfluency and predicting nonliteral meanings

2. Study human reading behavior for written disfluency and literal/nonliteral meanings

materials: grammatically correct speech-like sentences

materials: grammatically correct speech-like sentences

I'm sure negative peer pressure leads to mostly dumb decisions. I know from experience.

I'm sure negative peer pressure leads to mostly wise decisions. I know from experience.

materials: grammatically correct speech-like sentences

literal	I'm sure negative peer pressure leads to mostly dumb decisions. I know from experience.
nonliteral	I'm sure negative peer pressure leads to mostly wise decisions. I know from experience.

materials: grammatically correct speech-like sentences

literal	I'm sure negative peer pressure leads to mostly dumb decisions. I know from experience.
nonliteral	I'm sure negative peer pressure leads to mostly wise decisions. I know from experience.

norming: each participant would only rate one version of each item

How sarcastic do you think the author of this sentence was being?

Not sarcastic at all	Probably not sarcastic	Possibly not sarcastic	Might or might not be sarcastic	Probably sarcastic	Very probably sarcastic	Definitely sarcastic

self-paced reading 24 items

BERTweet / cloze test / eye-tracking reading 48 counterbalanced items

self-paced reading

24 items

merry (literal)/feral (nonliteral)

stupid (literal)/smart (nonliteral)

BERTweet / cloze test / eye-tracking reading

48 counterbalanced items

self-paced reading

`24 items`

`merry (literal)/feral (nonliteral)`
`stupid (literal)/smart (nonliteral)`

BERTweet / cloze test / eye-tracking reading

`48 counterbalanced items`

> Literal and nonliteral readings of each word are counterbalanced.

self-paced reading

24 items

merry (literal)/feral (nonliteral)
stupid (literal)/smart (nonliteral)

BERTweet / cloze test / eye-tracking reading

48 counterbalanced items

> Literal and nonliteral readings of each word are counterbalanced.

> The literal and nonliteral words have the same number of characters.

self-paced reading

24 items

merry (literal)/feral (nonliteral)
stupid (literal)/smart (nonliteral)

BERTweet / cloze test / eye-tracking reading

48 counterbalanced items

> Literal and nonliteral readings of each word are counterbalanced.

> The literal and nonliteral words have the same number of characters.

merry (literal)/feral (nonliteral)
feral (literal)/merry (nonliteral)
stupid (literal)/brainy (nonliteral)
brainy (literal)/stupid (nonliteral)

self-paced reading

24 items

merry (literal)/**feral** (nonliteral)
stupid (literal)/**smart** (nonliteral)

BERTweet / cloze test / eye-tracking reading

48 counterbalanced items

> Literal and nonliteral readings of each word are counterbalanced.

> The literal and nonliteral words have the same number of characters.

merry (literal)/**feral** (nonliteral)
feral (literal)/**merry** (nonliteral)
stupid (literal)/**brainy** (nonliteral)
brainy (literal)/**stupid** (nonliteral)

> Commas are added before and after *um* and the items are longer.

… **be, um, merry when** …

materials in 4 conditions based on fluency and meaning

materials in 4 conditions based on fluency and meaning

literal
fluent

Sitting through an hour of sermon would make most children feral on any day. You can ask them.

nonliteral
fluent

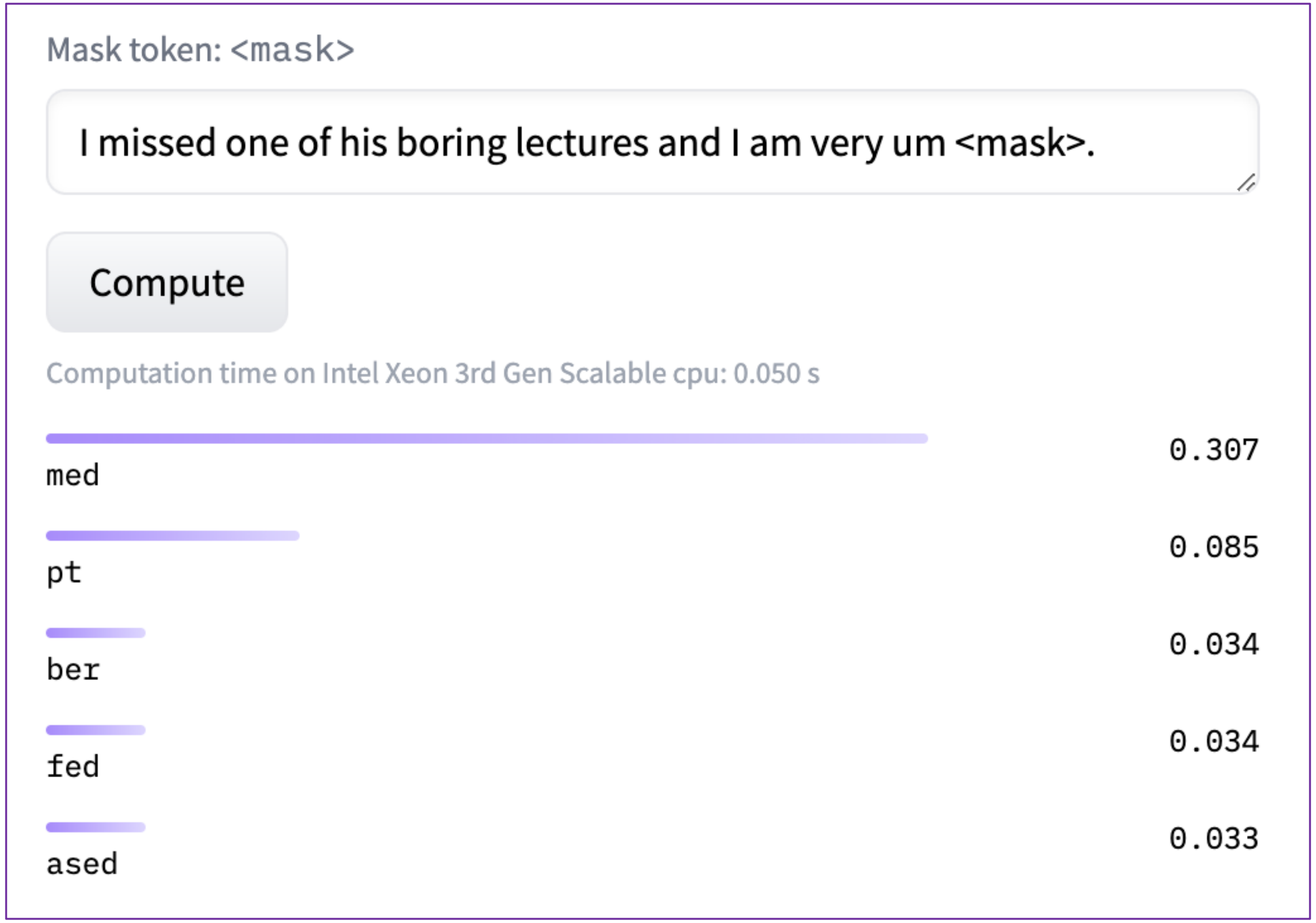
Sitting through an hour of sermon would make most children merry on any day. You can ask them.

materials in 4 conditions based on fluency and meaning

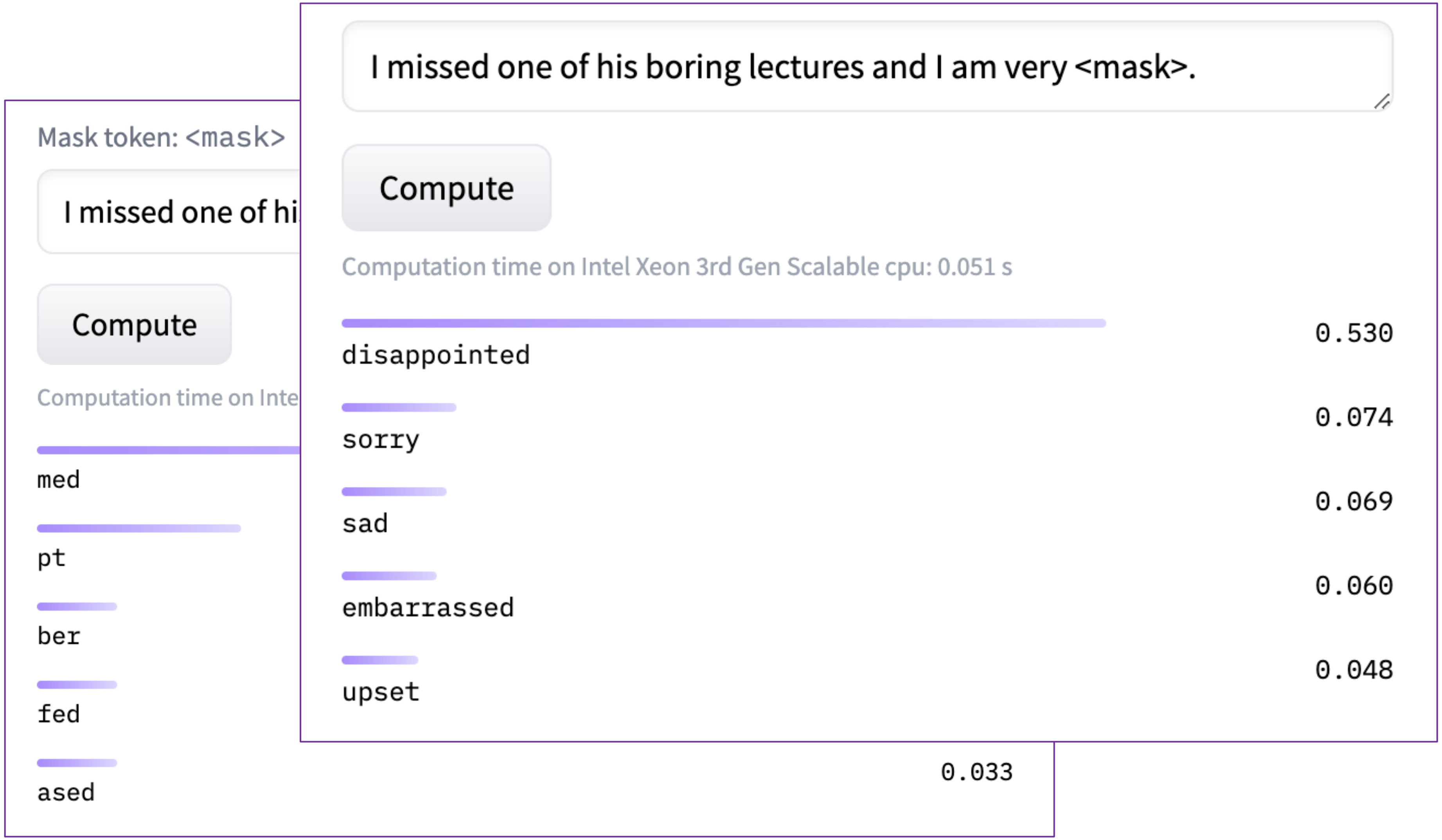
literal fluent	`Sitting through an hour of sermon would make most children feral on any day. You can ask them.`
literal disfluent	`Sitting through an hour of sermon would make most children, um, feral on any day. You can ask them.`
nonliteral fluent	`Sitting through an hour of sermon would make most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

1 Compare LM and human treatment of written disfluency and predicting nonliteral meanings

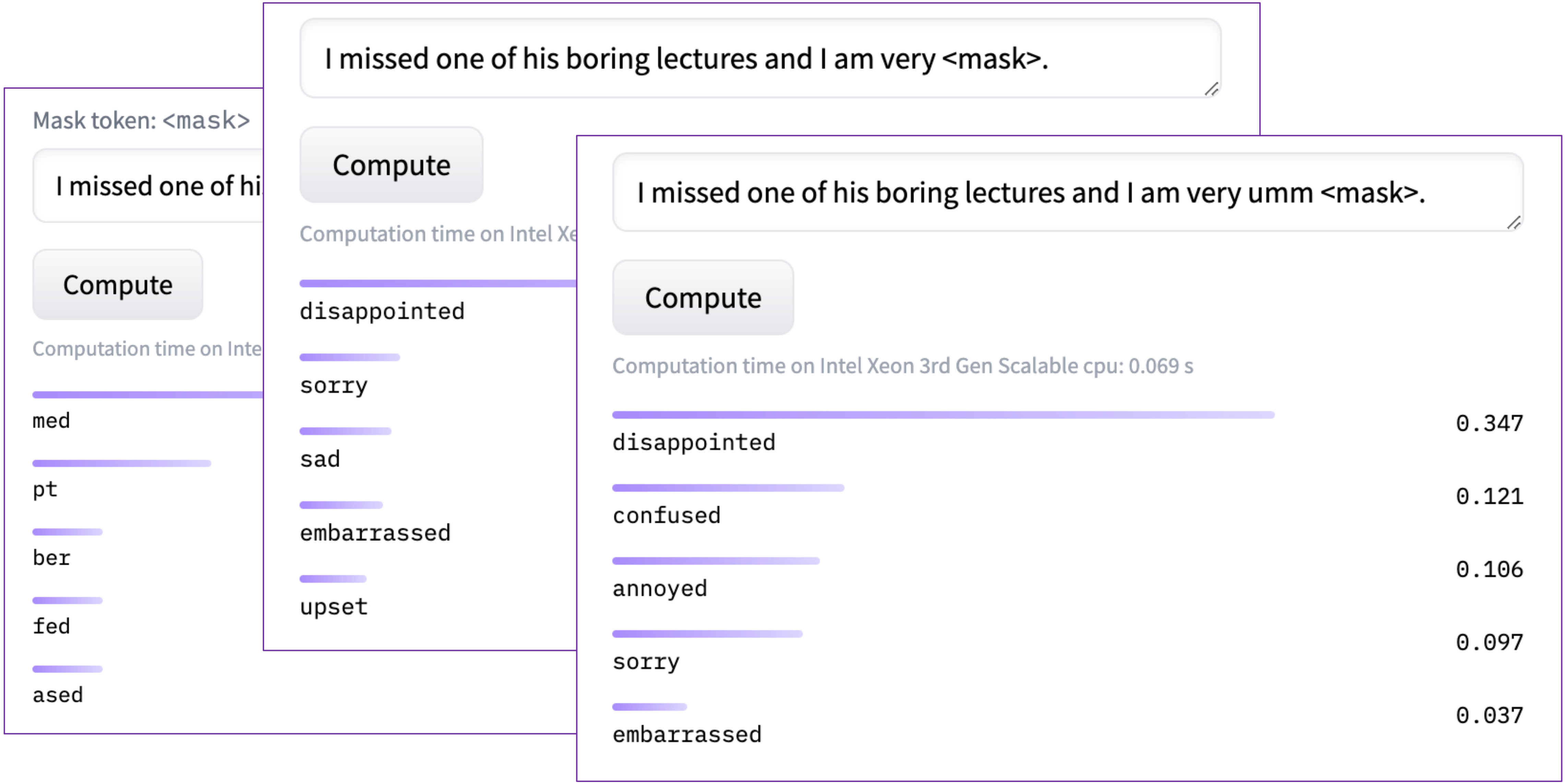
RoBERTa base:



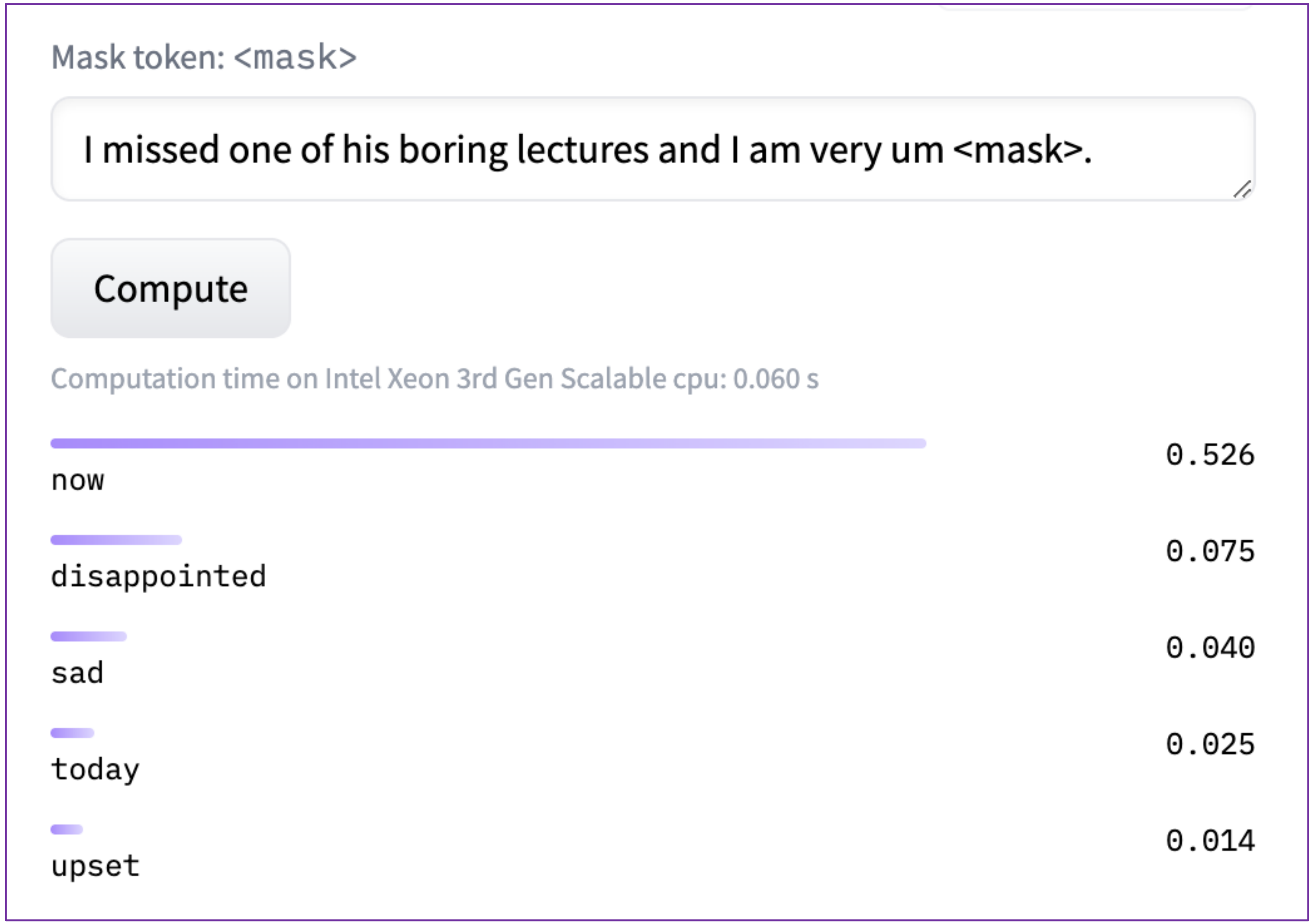
RoBERTa base:



RoBERTa base:

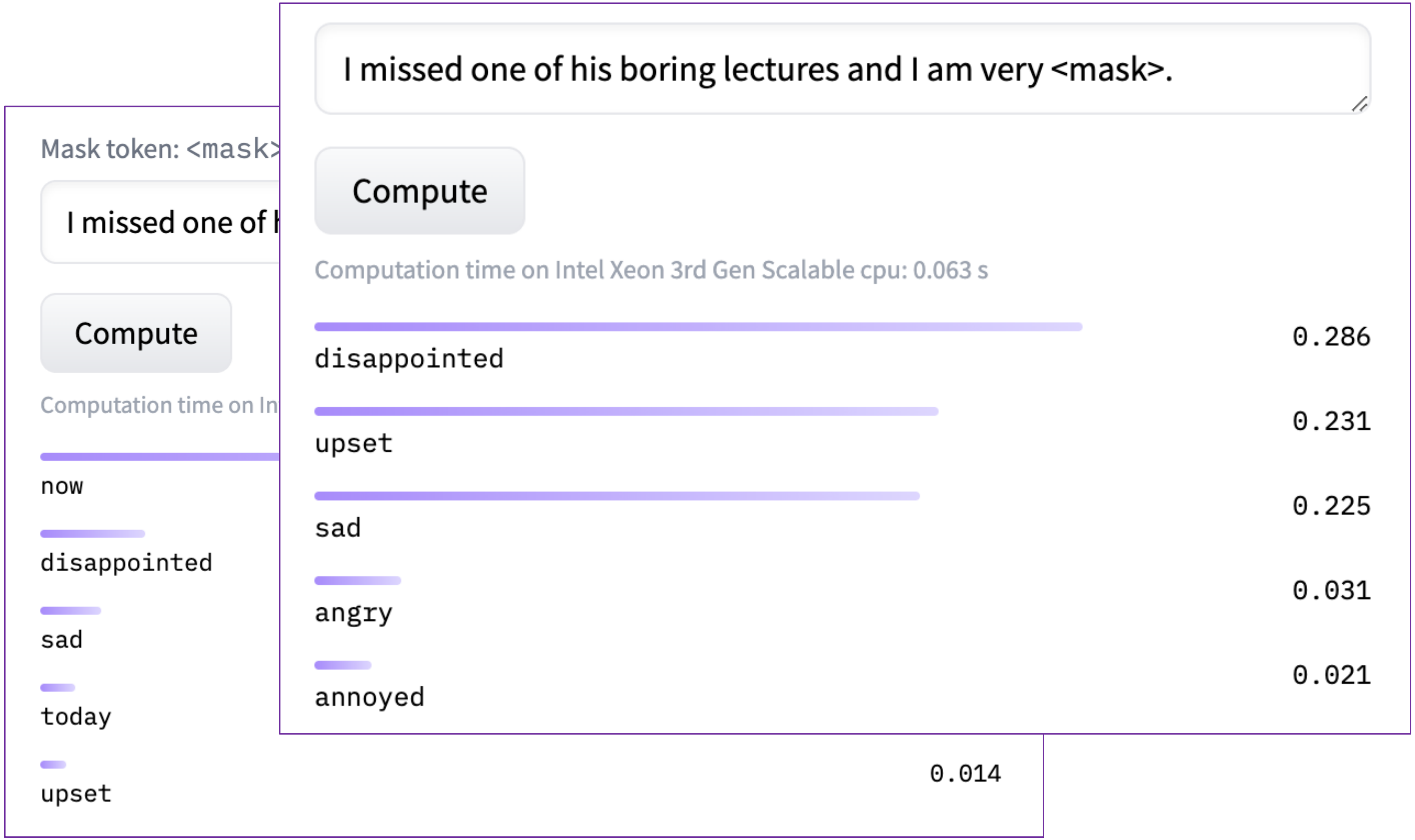


BERTweet:

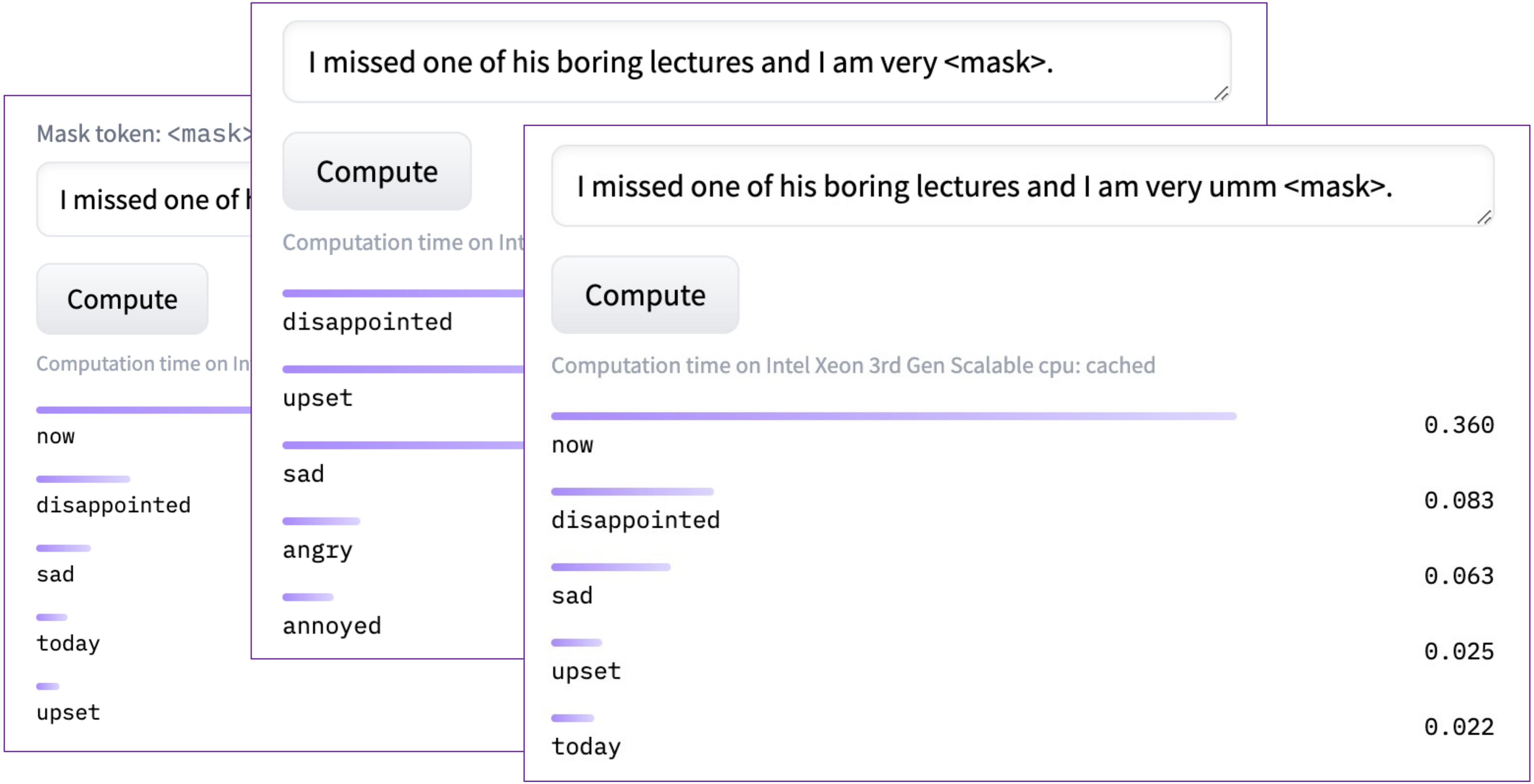


Mask token: <mask>
I missed one of his boring lectures and I am very um <mask>.
Compute
Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.060 s
now 0.526
disappointed 0.075
sad 0.040
today 0.025
upset 0.014

BERTweet:


Mask token: <mask>
I missed one of h
Compute
Computation time on In
now
disappointed
sad
today
upset
0.014
I missed one of his boring lectures and I am very <mask>.
Compute
Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.063 s
disappointed
0.286
upset
0.231
sad
0.225
angry
0.031
annoyed
0.021

BERTweet:


Mask token: <mask>
I missed one of h
Compute
Computation time on In
now
disappointed
sad
today
upset
I missed one of his boring lectures and I am very <mask>.
Compute
Computation time on Int
disappointed
upset
sad
angry
annoyed
I missed one of his boring lectures and I am very umm <mask>.
Compute
Computation time on Intel Xeon 3rd Gen Scalable cpu: cached
now 0.360
disappointed 0.083
sad 0.063
upset 0.025
today 0.022

masked language modeling: BERTweet

masked language modeling: BERTweet

fluent
~~stupid~~

Keep speaking nonsense and people will think you are <mask> at some point. I'm telling you.

disfluent
~~brainy~~

Keep speaking nonsense and people will think you are, um, <mask> at some point. I'm telling you.

masked language modeling: BERTweet

fluent
~~stupid~~

Keep speaking nonsense and people will think you are <mask> at some point. I'm telling you.

disfluent
~~brainy~~

Keep speaking nonsense and people will think you are, um, <mask> at some point. I'm telling you.

1. See what's happening in the top 10 predictions. (Does it get nonsensical at some point?)

2. Compare the top one to the most frequent cloze completion token.

cloze completion: humans

160 participants: 80 in each condition

First word that comes to mind – only one word, no hyphens

cloze completion: humans

160 participants: 80 in each condition

First word that comes to mind – only one word, no hyphens

1/48

Keep speaking nonsense and people will think you are, um, ______ at some point. I'm telling you.

disfluent

~~stupid~~

1/48

Keep speaking nonsense and people will think you are ______ at some point. I'm telling you.

Word count: 0/1

fluent

~~brainy~~

measures

Latent Semantic Analysis (LSA) scores

measures

Latent Semantic Analysis (LSA) scores

by utilizing pairwise comparisons using word2vec (Google News, 300 dimensions) for word embedding analysis

measures

Latent Semantic Analysis (LSA) scores

by utilizing pairwise comparisons using word2vec (Google News, 300 dimensions) for word embedding analysis

similarity score between cloze completions and BERTweet predictions for each item

measures

Latent Semantic Analysis (LSA) scores

by utilizing pairwise comparisons using word2vec (Google News, 300 dimensions) for word embedding analysis

similarity score between cloze completions and BERTweet predictions for each item

by multiplying the **number of identical cloze completions** by the **BERTweet confidence scores** and then by the **LSA cosine similarity** between words

Well, blue whales are an endangered species. So, I'd say hunting them is a really <mask> choice environmentally speaking.

poor (literal)/good (nonliteral)

Well, blue whales are an endangered species. So, I'd say hunting them is a really <mask> choice environmentally speaking.

poor (literal)/good (nonliteral)

The most popular cloze completion word chosen by 30 participants was ***bad***.

```
Well, blue whales are an endangered species. So, I'd say hunting them is
a really <mask> choice environmentally speaking.
```

```
poor (literal)/good (nonliteral)
```

The most popular cloze completion word chosen by 30 participants was ***bad***.

The highest ranked LM completion was ***good***, which had a confidence rating of 0.333.

Well, blue whales are an endangered species. So, I'd say hunting them is a really <mask> choice environmentally speaking.

poor (literal)/good (nonliteral)

The most popular cloze completion word chosen by 30 participants was ***bad***.

The highest ranked LM completion was ***good***, which had a confidence rating of 0.333.

The word2vec similarity score between ***bad*** and ***good*** was 0.719.

```
Well, blue whales are an endangered species. So, I'd say hunting them is a really <mask> choice environmentally speaking.
```

```
poor (literal)/good (nonliteral)
```

The most popular cloze completion word chosen by 30 participants was ***bad***.

The highest ranked LM completion was ***good***, which had a confidence rating of 0.333.

The word2vec similarity score between ***bad*** and ***good*** was 0.719.

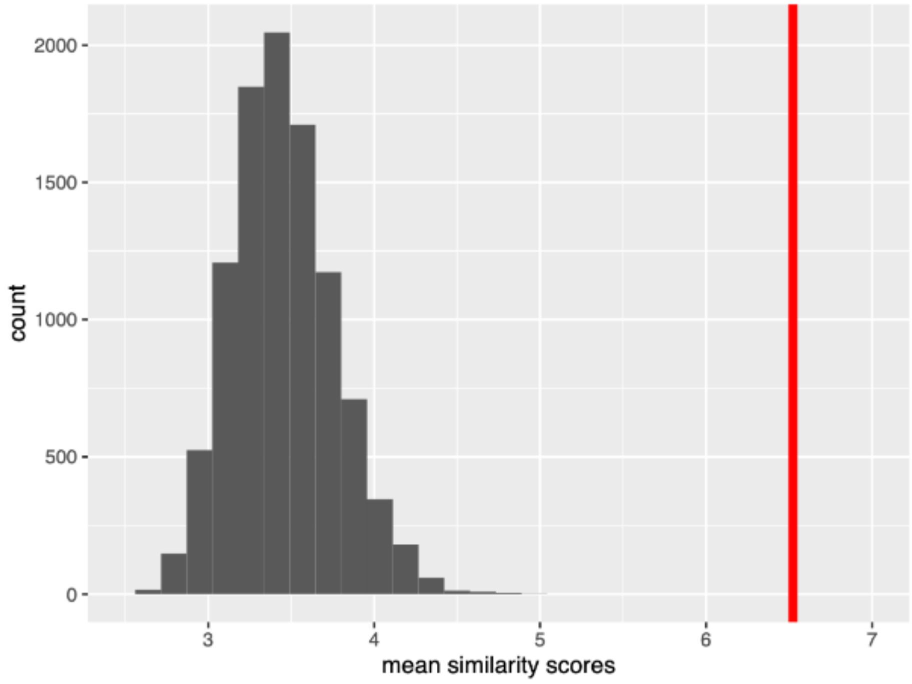
overall score:

$$\boxed{30 \times 0.333 \times 0.719 = 7.183}$$

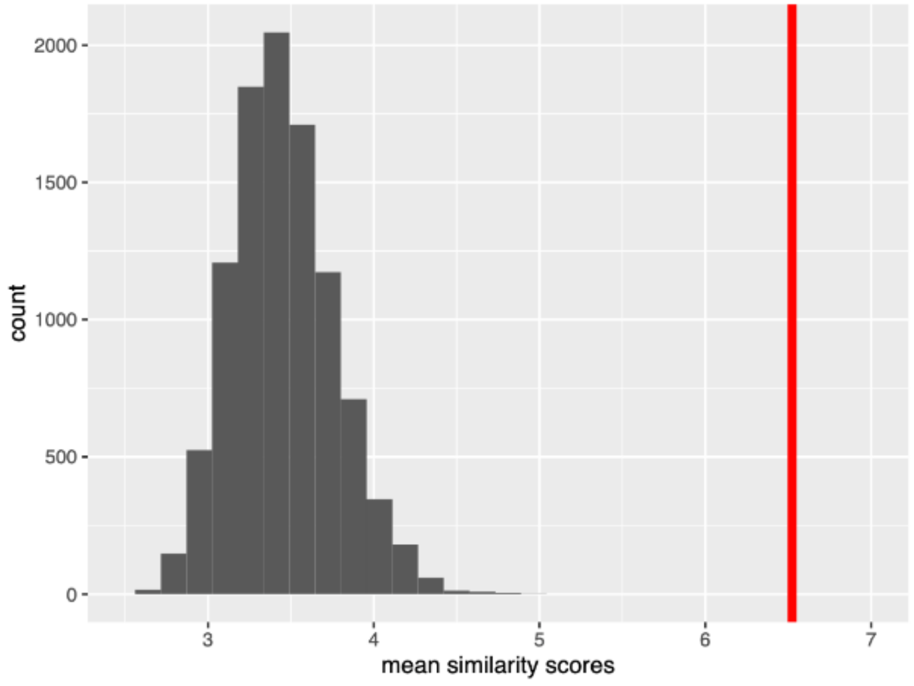
The mean similarity score between BERTweet and human cloze completions was **6.52**.

The mean similarity score between BERTweet and human cloze completions was **6.52**.

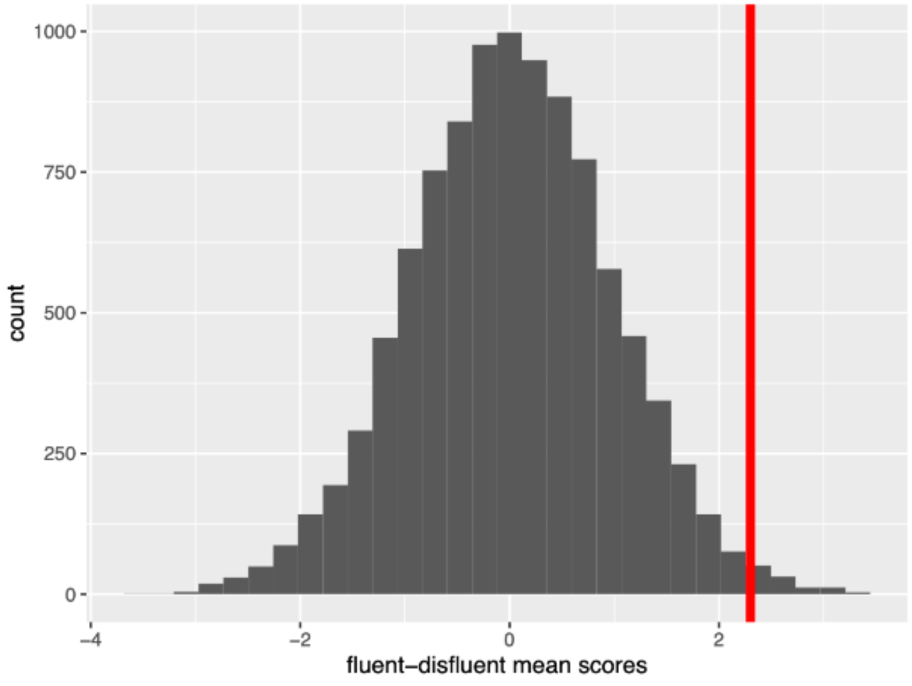
To assess BERTweet's performance against chance, we ran 10,000 permutations of the scores and recalculated the mean similarity.



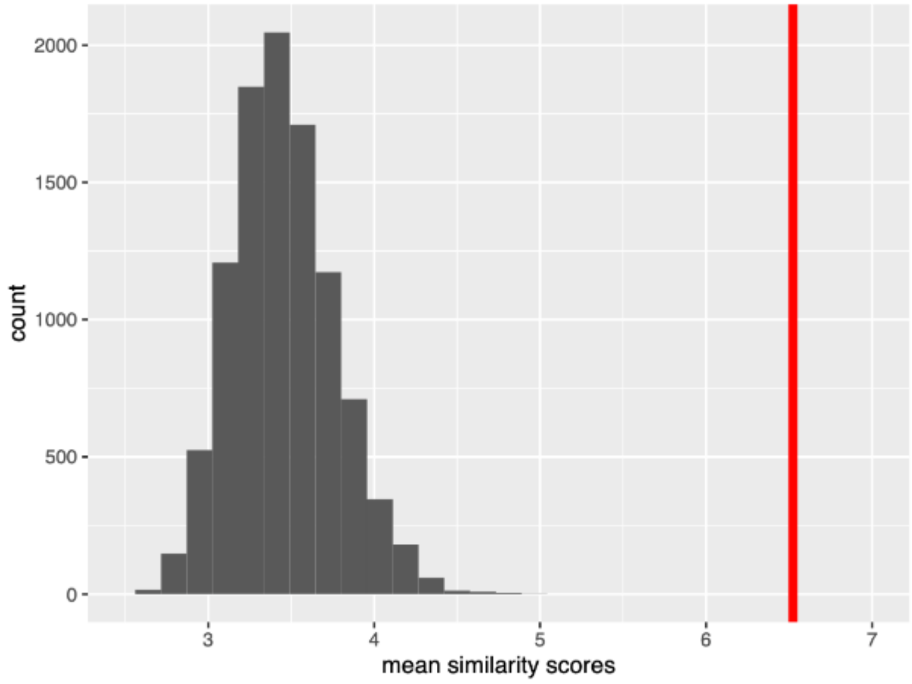
Recalculated mean similarity scores after 10k permutations, with the red vertical line indicating the mean similarity score of 6.52.

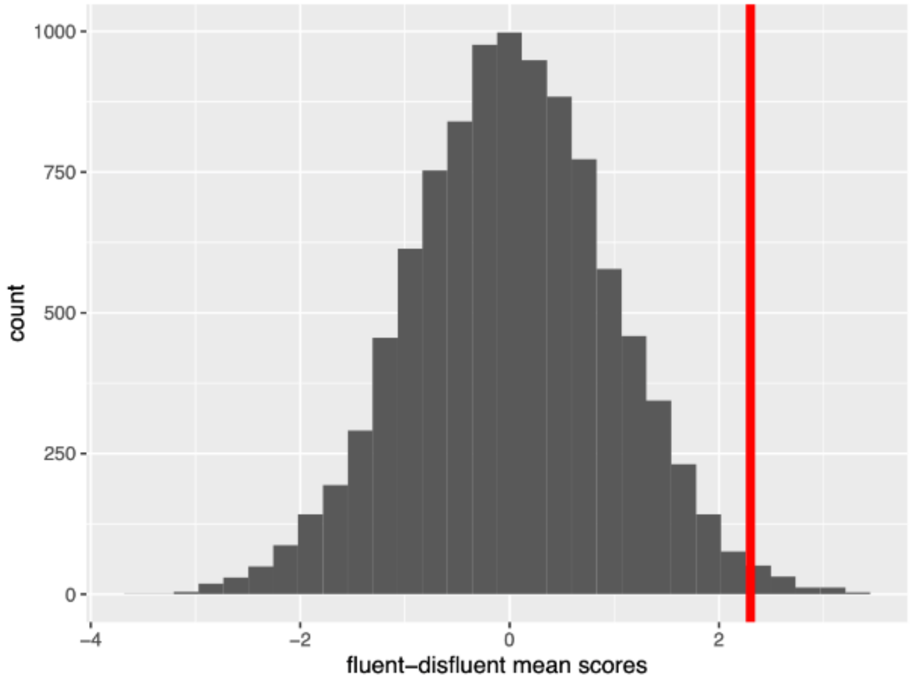


Recalculated mean similarity scores after 10k permutations, with the red vertical line indicating the mean similarity score of 6.52.



Recalculated fluent-disfluent mean similarity scores after 10k permutations, with the red vertical line indicating the difference in similarity score of 2.30.





Recalculated mean similarity scores after 10k permutations, with the red vertical line indicating the mean similarity score of 6.52.

Recalculated fluent-disfluent mean similarity scores after 10k permutations, with the red vertical line indicating the difference in similarity score of 2.30.

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

2 **Study human reading behavior for written disfluency and literal/nonliteral meanings**

reading behavior

reading behavior

1. Words compatible with a nonliteral/sarcastic reading of a sentence should be easier to read when preceded by *um* than when not preceded by *um.*

reading behavior

1. Words compatible with a nonliteral/sarcastic reading of a sentence should be easier to read when preceded by *um* than when not preceded by *um.*

...hunting blue whales is a really WISE move

...hunting blue whales is a really UM WISE move

reading behavior

1. Words compatible with a nonliteral/sarcastic reading of a sentence should be easier to read when preceded by *um* than when not preceded by *um.*

> *...hunting blue whales is a really WISE move*
>
> *...hunting blue whales is a really UM WISE move*

2. Words compatible with a literal reading of a sentence might be harder to read when preceded by *um* than when not preceded by *um.*

reading behavior

1. Words compatible with a nonliteral/sarcastic reading of a sentence should be easier to read when preceded by *um* than when not preceded by *um.*

…hunting blue whales is a really WISE move

…hunting blue whales is a really UM WISE move

2. Words compatible with a literal reading of a sentence might be harder to read when preceded by *um* than when not preceded by *um.*

…hunting blue whales is a really BAD move

…hunting blue whales is a really UM BAD move

longer reading times and/or more regressions for **fluent literal** items than **disfluent nonliteral** ones

There would be an interaction between fluency and meaning:

i.e., *um* would signal a shift toward nonliteral meaning.

word-by-word self-paced reading

word-by-word self-paced reading

literal
fluent

I'm sure negative peer pressure leads to mostly idiotic decisions.

sarcastic
fluent

I'm sure negative peer pressure leads to mostly clever decisions.

word-by-word self-paced reading

literal fluent	I'm sure negative peer pressure leads to mostly idiotic decisions.
literal disfluent	I'm sure negative peer pressure leads to mostly um idiotic decisions.
sarcastic fluent	I'm sure negative peer pressure leads to mostly clever decisions.
sarcastic disfluent	I'm sure negative peer pressure leads to mostly um clever decisions.

word-by-word self-paced reading

P1	I'm sure negative peer pressure leads to mostly idiotic decisions.
P2	I'm sure negative peer pressure leads to mostly um idiotic decisions.
P3	I'm sure negative peer pressure leads to mostly clever decisions.
P4	I'm sure negative peer pressure leads to mostly um clever decisions.

Follows the structure of an ILS Labs moving window experiment using jsPsych.

word-by-word self-paced reading

101 participants: L1-English, UK-based, and non-dyslexic

26 items: 2 practice items and 24 experimental items

6 items in each experimental condition

8 attention checks

word-by-word self-paced reading

P2

I'm

word-by-word self-paced reading

P2 sure

word-by-word self-paced reading

P2 `negative`

word-by-word self-paced reading

P2

peer

word-by-word self-paced reading

P2

pressure

word-by-word self-paced reading

P2

leads

word-by-word self-paced reading

P2

to

word-by-word self-paced reading

P2

mostly

word-by-word self-paced reading

P2

um

word-by-word self-paced reading

P2

idiotic

word-by-word self-paced reading

P2

decisions.

word-by-word self-paced reading

target

P2 I'm sure negative peer pressure leads to mostly um idiotic decisions.

target + next

spill-over

reading times in milliseconds

linear mixed-effects models using log-transformed reading times

linear mixed-effects models using log-transformed reading times

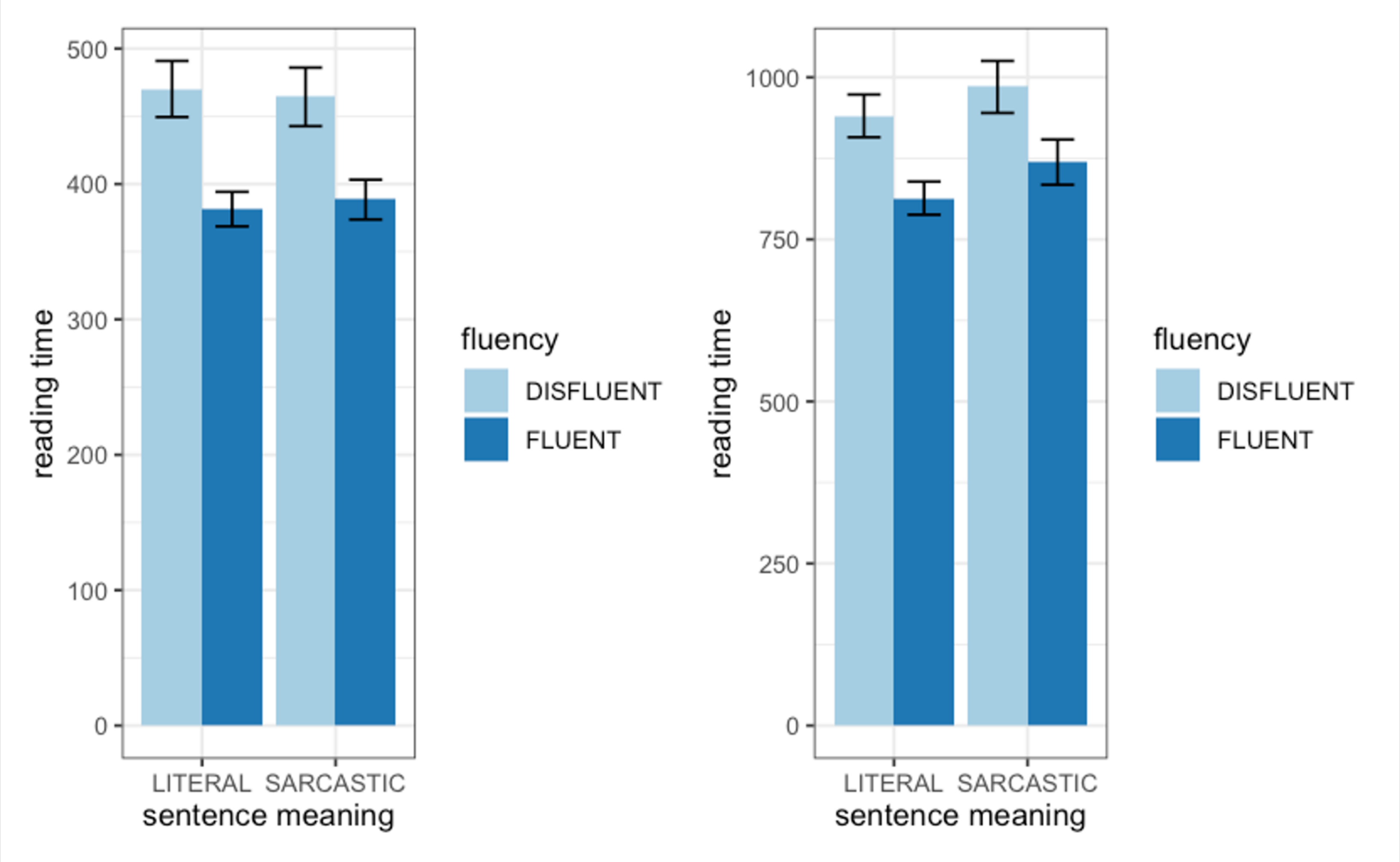
☒ When preceded by um, the nonliteral sentences were not faster to read.

linear mixed-effects models using log-transformed reading times

- ☒ When preceded by um, the nonliteral sentences were not faster to read.
- ☑ Literal sentences were faster to read than nonliteral sentences.
- ☑ Disfluent sentences were slower to read than fluent sentences.

word-by-word self-paced reading of 24 items by 99 participants

reading times in milliseconds

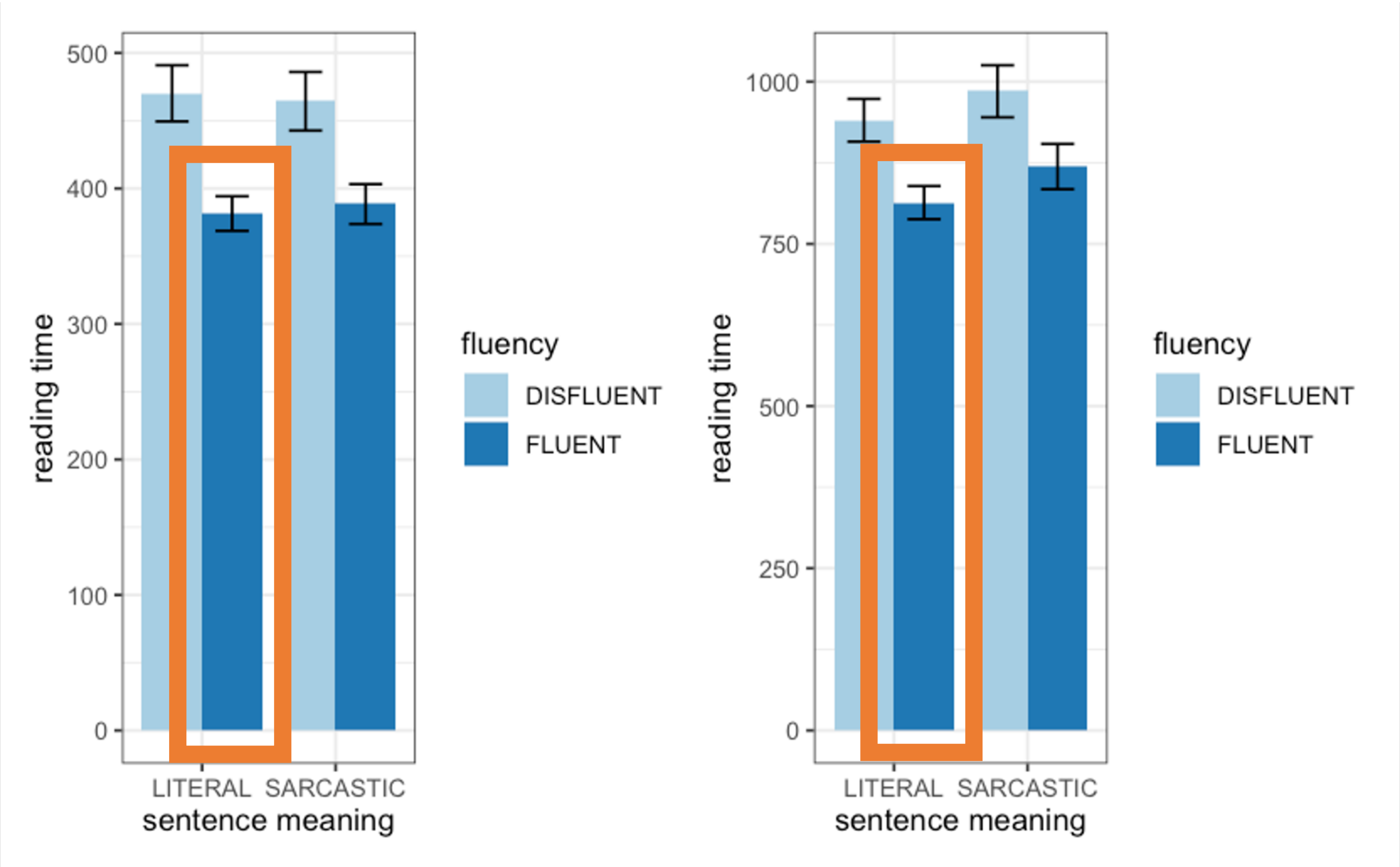


target region

target + next regions

word-by-word self-paced reading of 24 items by 99 participants

reading times in milliseconds



target region

target + next regions

The SPR experiment didn’t show that written disfluency indexes nonliteral meaning, at least, in the form of sarcasm.

But it did show that readers were sensitive to written *um*.

The SPR experiment didn't show that written disfluency indexes nonliteral meaning, at least, in the form of sarcasm.

But it did show that readers were sensitive to written *um*.

It could be that the artificial segmentation needed for self-paced reading disrupted the rhythm with which readers might have read the experimental sentences, reducing any interruption effect that the traditionally spoken element *um* might have had in writing.

eye-tracking reading

eye-tracking reading

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.

eye-tracking reading

P1 Sitting through an hour of sermon would make most children feral on any day. You can ask them.

P2 Sitting through an hour of sermon would make most children, um, feral on any day. You can ask them.

P3 Sitting through an hour of sermon would make most children merry on any day. You can ask them.

P4 Sitting through an hour of sermon would make most children, um, merry on any day. You can ask them.

Uses Experiment Builder to present items on an EyeLink 1000 Plus tracker for in-person data collection.

eye-tracking reading

60 participants: neurotypical, L1-English, 18-34 years old, no reported reading disorders, normal/surgically-corrected-to-normal vision

152 items: 2 practice items, 48 experimental items, 102 filler items

12 items in each experimental condition

32 attention checks: 16 for experimental items and 16 for filler items

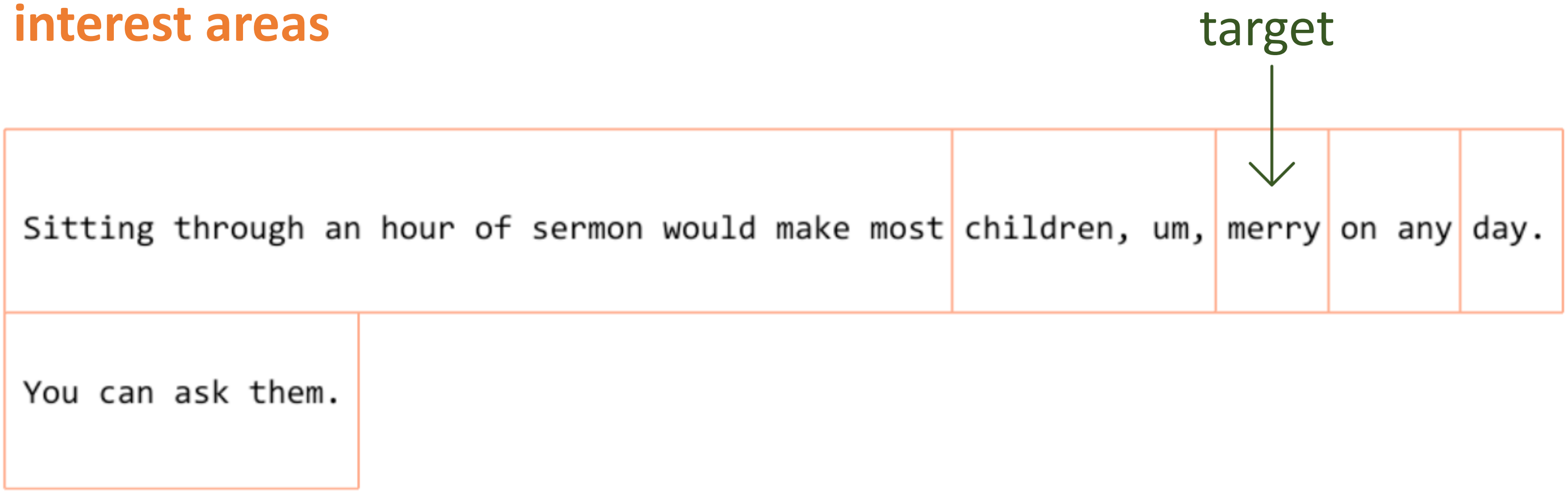
eye-tracking reading

interest areas

Sitting through an hour of sermon would make most	children, um,	merry	on any	day.
You can ask them.				

eye-tracking reading

interest areas



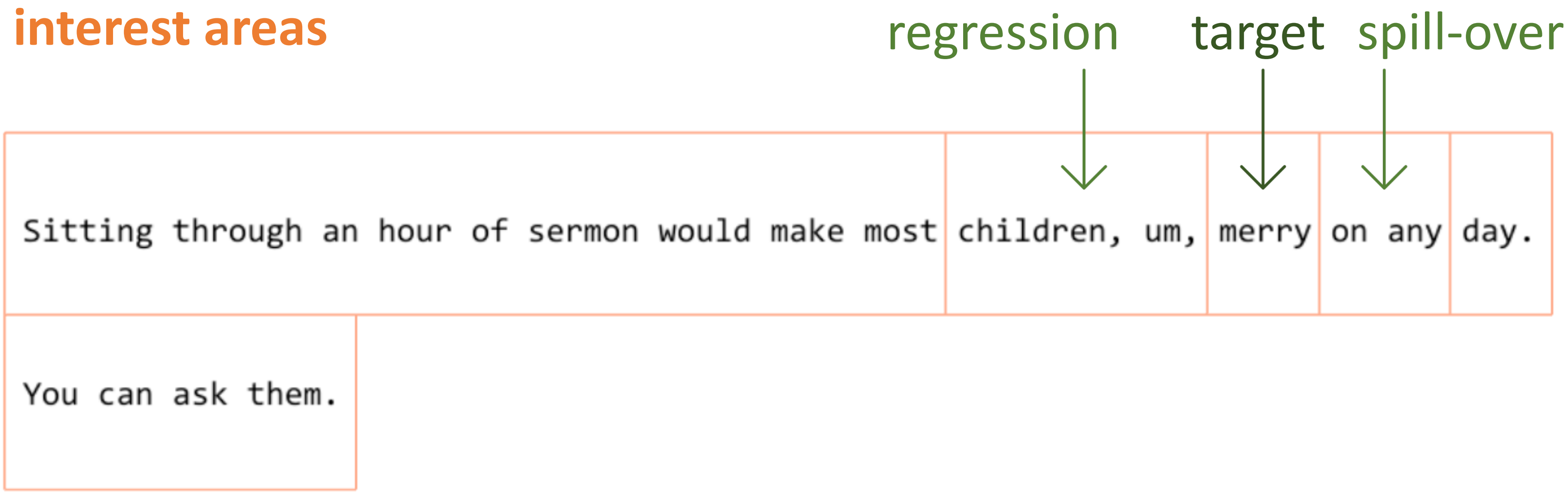
eye-tracking reading

interest areas

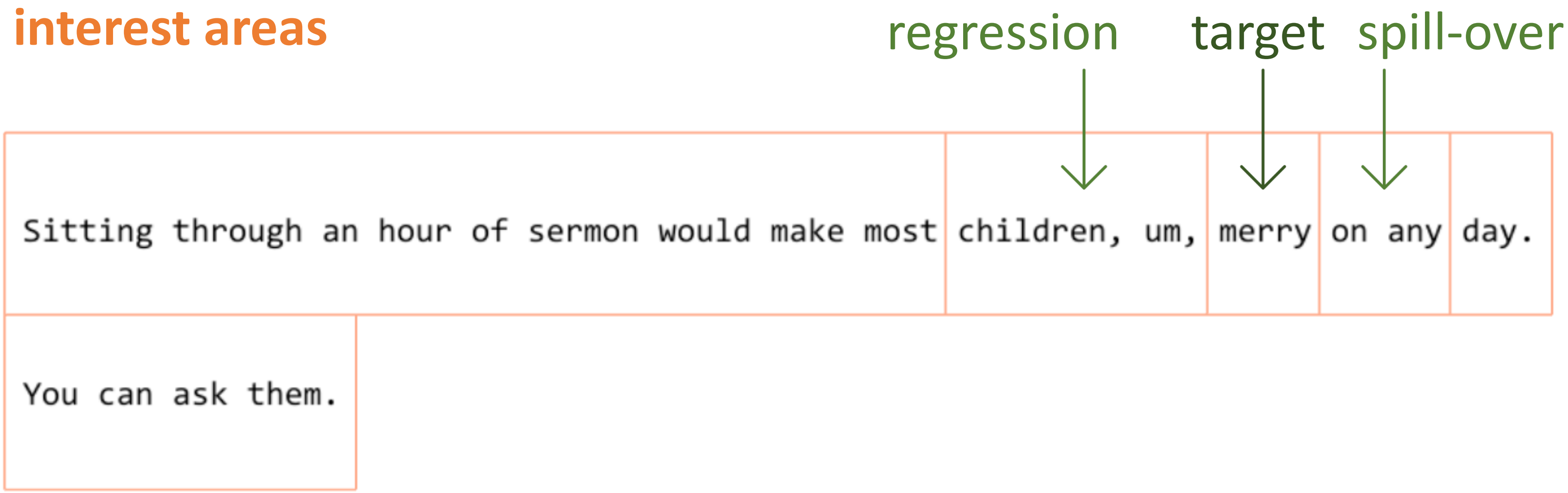
target spill-over

Sitting through an hour of sermon would make most	children, um,	merry	on any	day.
You can ask them.				

eye-tracking reading



eye-tracking reading



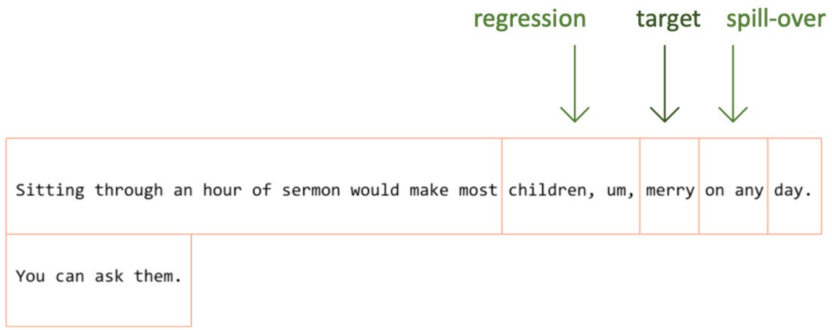
reading times in milliseconds

eye-tracking reading

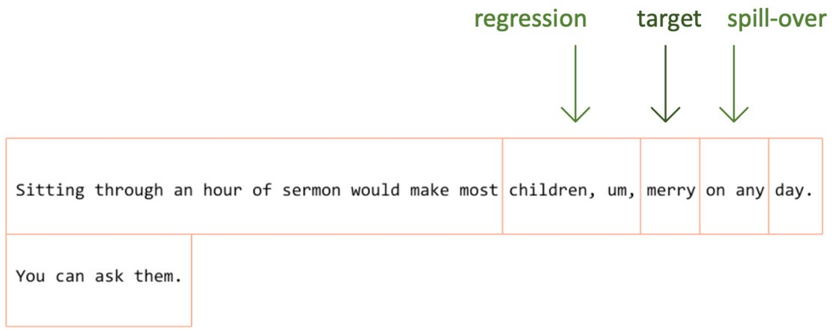
fixations
right eye

eye-tracking reading

measures:



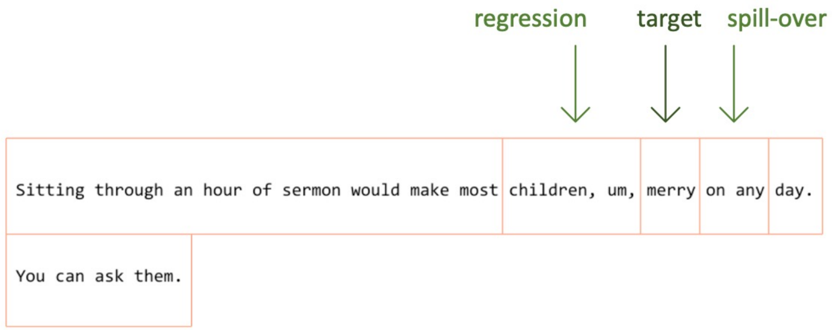
eye-tracking reading



measures:

1. **regression path time** (go-past time) which is the summed fixation duration from when the current interest area is first fixated until the eyes enter a later interest area

eye-tracking reading

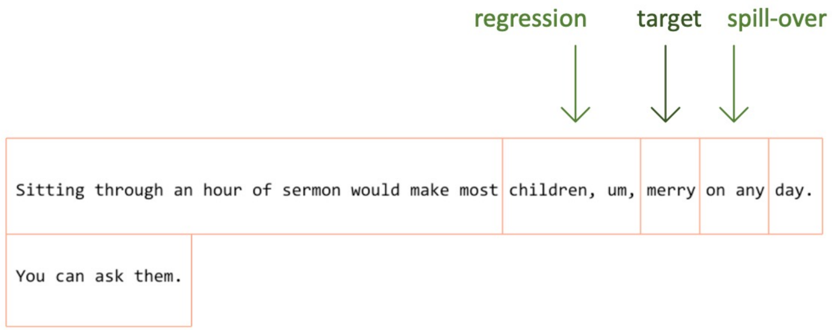


measures:

1. **regression path time** (go-past time) which is the summed fixation duration from when the current interest area is first fixated until the eyes enter a later interest area

2. **first pass time** which is the sum of the duration of all fixations before the interest area is exited for the first time

eye-tracking reading



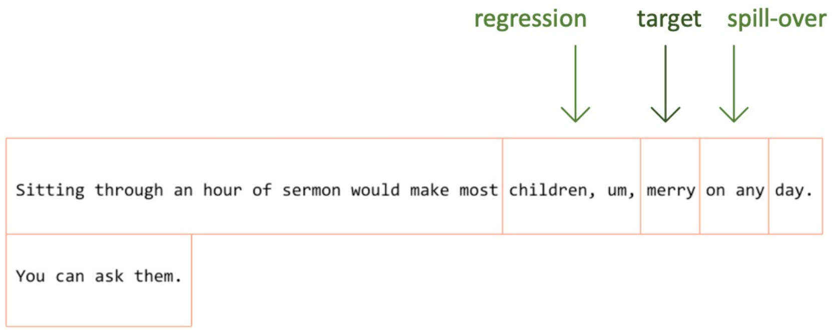
measures:

1. **regression path time** (go-past time) which is the summed fixation duration from when the current interest area is first fixated until the eyes enter a later interest area

2. **first pass time** which is the sum of the duration of all fixations before the interest area is exited for the first time

3. **total dwell time** which is the summation of the duration across all fixations on the current interest area

eye-tracking reading



measures:

1. **regression path time** (go-past time) which is the summed fixation duration from when the current interest area is first fixated until the eyes enter a later interest area

2. **first pass time** which is the sum of the duration of all fixations before the interest area is exited for the first time

3. **total dwell time** which is the summation of the duration across all fixations on the current interest area

4. **first pass regressions out** (target and next regions) indicating whether regression(s) were made from the current interest area to the earlier interest area prior to leaving the interest area in a forward direction

linear mixed-effects models using log-transformed reading times

linear mixed-effects models using log-transformed reading times

regression path time:

(go-past time) the summed fixation duration from when the current interest area is first fixated until the eyes enter a later interest area

target interest area – the effect of **meaning** ($\beta = 0.06$, $SE = 0.02$, $p = .01$)
target + next interest areas – the effect of **meaning** ($\beta = 0.10$, $SE = 0.03$, $p < .001$)

Literal words were faster to read than nonliteral ones.

linear mixed-effects models using log-transformed reading times

first pass time:

sum of the duration of all fixations before the interest area is exited for the first time

target interest area – the effect of **fluency** (β = -0.10, SE = 0.02, $p < .001$)

Fluent items were faster to read than disfluent ones.

linear mixed-effects models using log-transformed reading times

first pass time:

sum of the duration of all fixations before the interest area is exited for the first time

target interest area – the effect of **fluency** ($\beta = -0.10$, $SE = 0.02$, $p < .001$)

Fluent items were faster to read than disfluent ones.

target + next interest areas – the effect of **fluency** ($\beta = -0.07$, $SE = 0.01$, $p < .001$)
target + next interest areas – the effect of **meaning** ($\beta = 0.06$, $SE = 0.02$, $p = .004$)

Fluent items were faster to read than disfluent ones.

Literal words were faster to read than nonliteral ones.

linear mixed-effects models using log-transformed reading times

total dwell time: *the summation of the duration across all fixations on the current interest area*

target interest area – the effect of **fluency** ($\beta = -0.10$, $SE = 0.02$, $p < .001$)
target interest area – the effect of **meaning** ($\beta = 0.10$, $SE = 0.03$, $p < .001$)

Fluent items were faster to read than disfluent ones.

Literal words were faster to read than nonliteral ones.

linear mixed-effects models using log-transformed reading times

total dwell time: *the summation of the duration across all fixations on the current interest area*

target interest area – the effect of **fluency** ($\beta = -0.10$, $SE = 0.02$, $p < .001$)
target interest area – the effect of **meaning** ($\beta = 0.10$, $SE = 0.03$, $p < .001$)

Fluent items were faster to read than disfluent ones.

Literal words were faster to read than nonliteral ones.

target + next interest areas – the effect of **fluency** ($\beta = -0.07$, $SE = 0.02$, $p < .001$)
target + next interest areas – the effect of **meaning** ($\beta = 0.11$, $SE = 0.03$, $p < .001$)

Fluent items were faster to read than disfluent ones.

Literal words were faster to read than nonliteral ones.

logistic mixed-effects models

first pass regression out:

whether regression(s) were made from the current interest area to the earlier interest area prior to leaving the interest area in a forward direction

target interest area – the effect of **fluency** ($\beta = 1.01$, $SE = 0.16$, $p < .001$)

Regressions were more likely to be made following a fixation on the target word when the items were fluent.

There was **no interaction** between **fluency** and **meaning**:

um **didn't signal nonliteral meaning**.

There was **no interaction** between **fluency** and **meaning**:

um **didn't signal nonliteral meaning.**

But we found the effects of **fluency** and **meaning** on reading:

Literal words were read faster, and disfluency slowed reading.

wrap-up

wrap-up

Although BERTweet made human-like predictions, its performance was significantly worse when *um* was present.

wrap-up

Although BERTweet made human-like predictions, its performance was significantly worse when *um* was present.

Our reading experiments showed that readers were sensitive to fluency and to meaning.

wrap-up

Although BERTweet made human-like predictions, its performance was significantly worse when *um* was present.

Our reading experiments showed that readers were sensitive to fluency and to meaning.

BERTweet's difficulty with written disfluencies may be due to training on filtered data that excludes disfluencies.

wrap-up

Although BERTweet made human-like predictions, its performance was significantly worse when *um* was present.

Our reading experiments showed that readers were sensitive to fluency and to meaning.

BERTweet's difficulty with written disfluencies may be due to training on filtered data that excludes disfluencies.

Maybe we should recognize the communicative value of disfluencies in online writing and not dismiss them as irrelevant noise.

limitations

limitations

Our experiments to date have investigated a specific disfluency in a specific language and context.

limitations

Our experiments to date have investigated a specific disfluency in a specific language and context.

Our results may have been influenced by the specific design and sample size.

limitations

Our experiments to date have investigated a specific disfluency in a specific language and context.

Our results may have been influenced by the specific design and sample size.

We have established that written disfluencies are worth investigating, with LMs as well as humans sensitive to their presence, but this study is just a starting point.

limitations

Our experiments to date have investigated a specific disfluency in a specific language and context.

Our results may have been influenced by the specific design and sample size.

We have established that written disfluencies are worth investigating, with LMs as well as humans sensitive to their presence, but this study is just a starting point.

To gain a more complete picture, we should pay attention to the naturalness of the stimuli used, and should aim to generalize the work to other languages and disfluencies.

ongoing and future work

ongoing and future work

Remaining work on BERTweet:

Remove commas.

Manipulate filler placements.

Further pre-train the model using our tweet data set.

ongoing and future work

Remaining work on BERTweet:

Remove commas.

Manipulate filler placements.

Further pre-train the model using our tweet data set.

Testing disfluencies on an LLM:

Do set parameters and/or training data affect model performance?

ongoing and future work

Remaining work on BERTweet:

Remove commas.

Manipulate filler placements.

Further pre-train the model using our tweet data set.

Testing disfluencies on an LLM:

Do set parameters and/or training data affect model performance?

Diversifying disfluency and nonliteral interpretation research:

What is the effect of neurodiversity and sociocultural factors?

questions and comments?

Get in touch!

Aida in

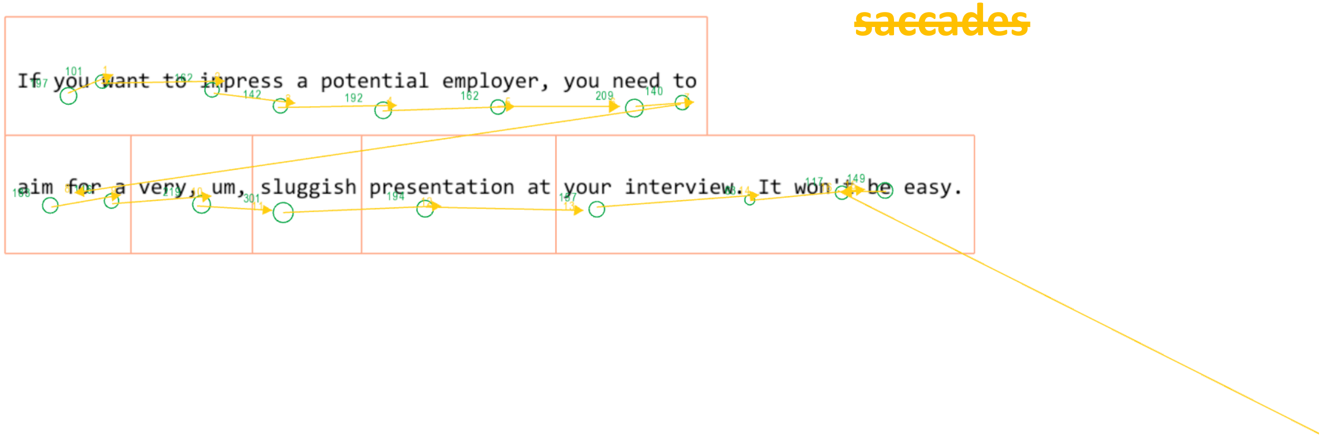
Martin https://martincorley.org/

Patrick patrick.sturt@ed.ac.uk

eye-tracking reading

fixations

~~saccades~~



eye-tracking reading

cleanup on Data Viewer:

eye-tracking reading

cleanup on Data Viewer:

automatic:

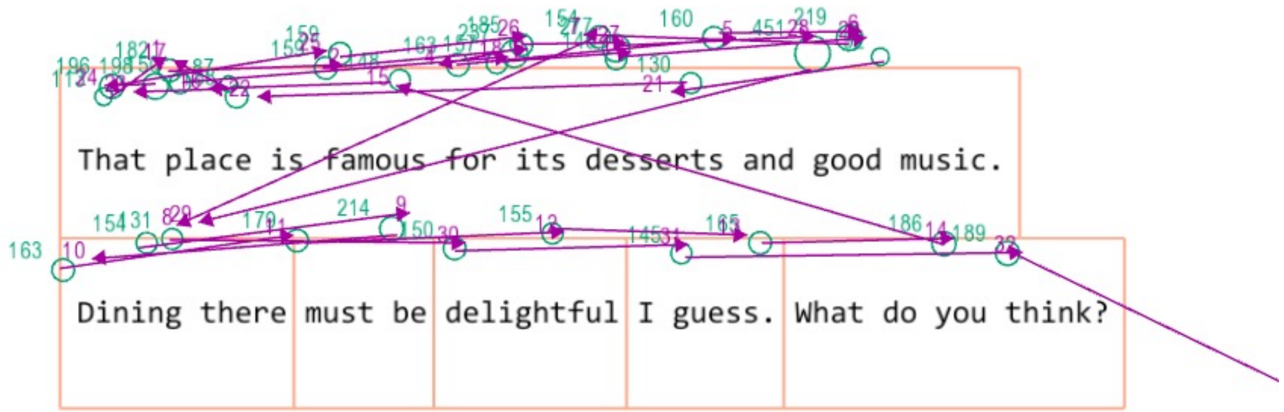
1. remove the filler trials
2. merge nearby fixations
3. remove fixations less than 80 milliseconds

eye-tracking reading

cleanup on Data Viewer:

automatic:

1. remove the filler trials
2. merge nearby fixations
3. remove fixations less than 80 milliseconds



manual:

1. align the fixations vertically within the preassigned interest area bounds
2. monitor the number of horizontally misaligned trials for each participant for removal

Previous study*

> We looked into how users were writing **um**, **uh**, **hmm**, **erm**, and **er** on Twitter.

> Participants rated 36 randomly selected tweets *with* and *without* **um** and **hmm** for their sarcastic tone, offensiveness, language formality, and the emotions associated with them.

With **um** and **hmm**,

- **sarcasm** scores were **slightly**, although not significantly, **higher**.

- tweets were perceived as **more offensive**, **less formal**, and **more surprising**.

*Tarighat, F. S., Magdy, W., & Corley, M. (2022). *Understanding Fillers May Facilitate Automatic Sarcasm Comprehension: A Structural Analysis of Twitter Data and a Participant Study*. Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue, Dublin.