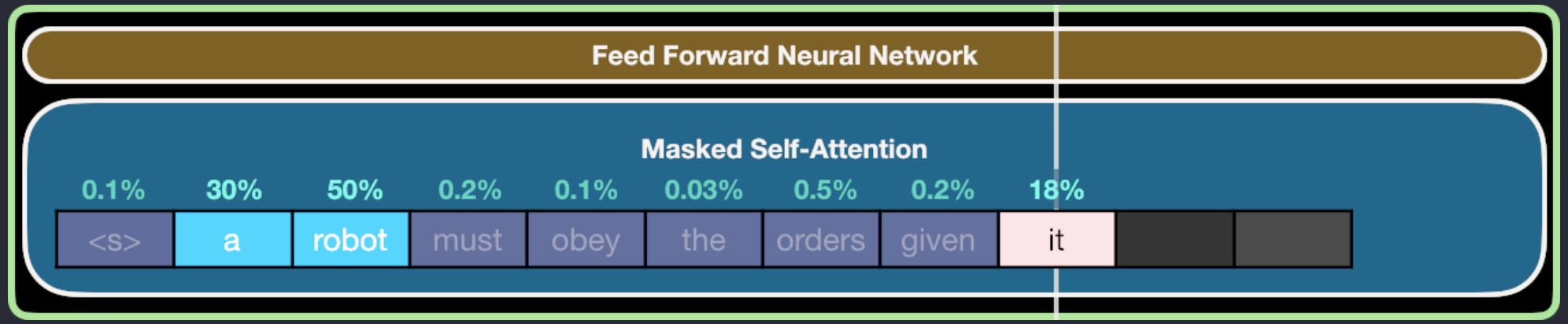
LLM Surprisal ~ Human Reading Time

Tobias Hoffmann, Lars Konieczny, Henrik Lorenzen, Mateo Cortés Lafourcade

Large Language Models

- Task: Given words so far, predict next word
  - e.g. prompt: *"The cat eats the "*, prediction: *"mouse"*
- State of the art: Attention mechanism[Vaswani et al. 2017]
  - Idea: Mix in *relevant* context tokens
  - Include large context without forgetfulness



LLM Surprisal

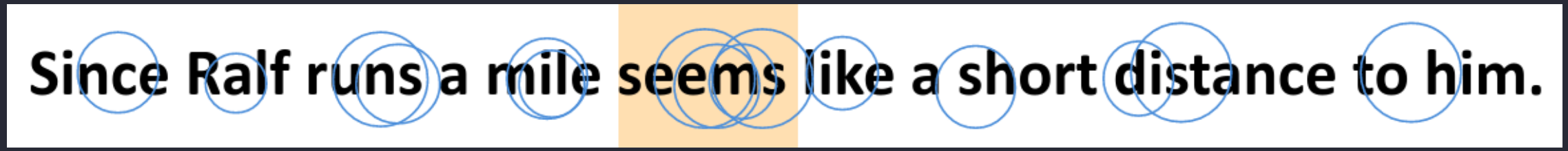
- Normally: Most likely next word?
- Instead: How likely is *this* next word?
  - I.e. $P(t_n|t_{n-1},\ldots,t_0)$
- *Surprisal* $:= -log_2 P(t_n|t_{n-1},\ldots,t_0)$
  - High surprisal $\Leftrightarrow$ Low probability
  - Between 0 and ∞

```
The cat eats the  mouse.
-   8.8 7.2  1.7  4.1

The cat eats the  car.
-   8.8 7.2  1.7  15.1
```

Human Reading Time

- How much time is spent per word while reading?
- Different metrics: *First Pass*, *Go Past*, etc.
- Generally: Predictable word ~ short reading time
  - We read over words that are predictable
  - We get stuck at words that are unexpected
- Measured in different ways, e.g. eye tracking



Local Syntactic Coherences

The coach chided the player tossed the frisbee by the opposing team.
The coach chided the player tossed the frisbee ***by*** the opposing team.

- Conflict between local and global parsing *Tabor et al. 2004*
- Increased reading time at subsequent word
- Not just syntactical, influenced by context
  - Including visual & textual *Konieczny et al. 2009, Müller et al. 2019*

Context influences LSC-Effect

Mit **(langweiligen Anekdoten | spannenden Geschichten)** überzieht der erste Redner sein Zeitlimit. Das Publikum hört ihm dabei **(gähnend | gespannt)** zu. Nach dem dreistündigen Vortrag hat er keine Energie mehr und übergibt an den nächsten Redner.

The first speaker exceeds his time limit with **(boring anecdotes | exciting stories)**. The audience listens to him **(yawning | attentively)**. After the three-hour talk, he has no more energy, and hands over to the next speaker.

Der nächste Redner ärgert sich über alle Maßen, als ihm der erste Redner **müde** das Publikum überlässt.

OR

Der nächste Redner ärgert sich über alle Maßen, als ihm der erste Redner **ermüdet** das Publikum ***überlässt***.

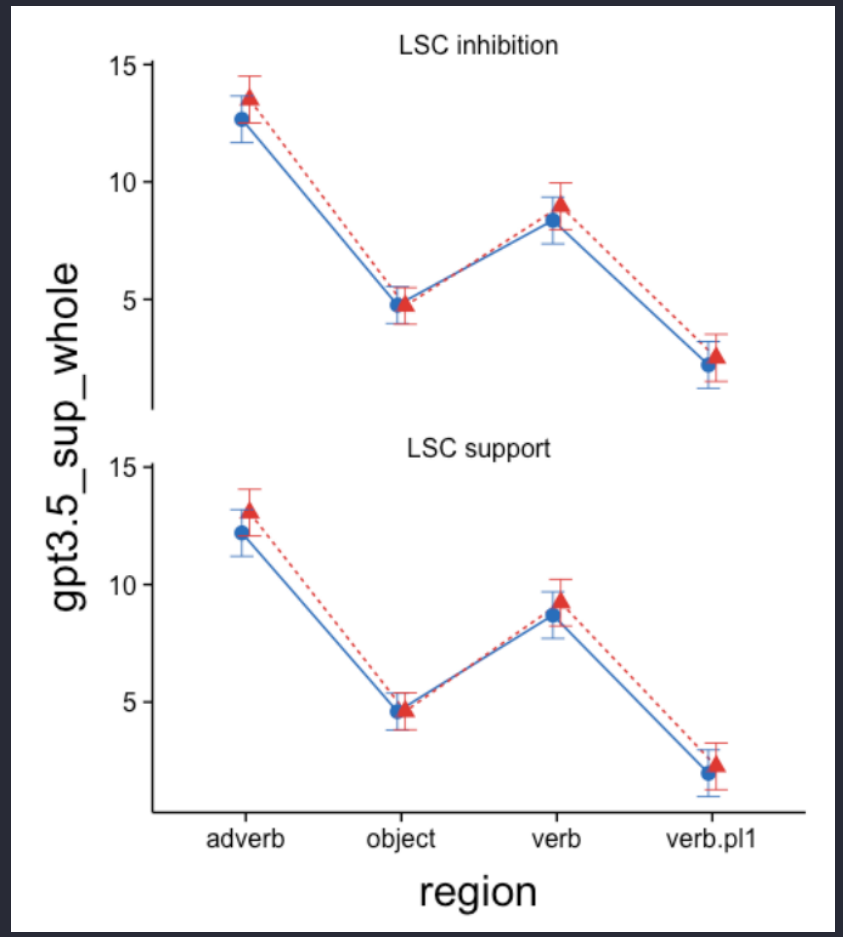
The next speaker is annoyed beyond measure, as the first speaker **(tiredly | tires)** leaves the audience to him.

Surprisal ~ Reading Time ?

- Dataset: short German texts *from Müller et al. 2019 & Müller 2019*
- Reading times via eye tracking
- GPT-3.5 surprisals via OpenAI API
- Yes! ✨

LSC-Effect in NNs?

- In RNNs: Yes *Konieczny 2005, Konieczny et al. 2009*
- In Transformers: No ✨
  - LSCs only mildly affect surprisal
  - No modulation due to context
  - Explanation (?):
    Attention mechanism has no intrinisic local preference



Conclusion

- LLM Surprisal predicts human reading times
- But LLMs aren't fooled by local syntactic coherences
- 💡 LLM Surprisal as quantitative research tool
- ⚠️ Limits of LLMs as models of human language processing

- **AMLAP Poster:** http://dx.doi.org/10.13140/RG.2.2.15402.39363/2
- **Slides at:** https://github.com/Garbaz/IICCSSS2023