Next-Word Prediction in Language Models and Humans

Tatsuki Kuribayashi (MBZUAI)

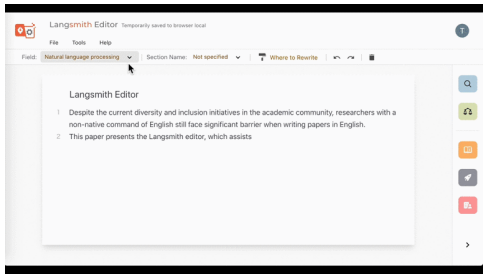
Research topics

Web page

Automated writing assistance

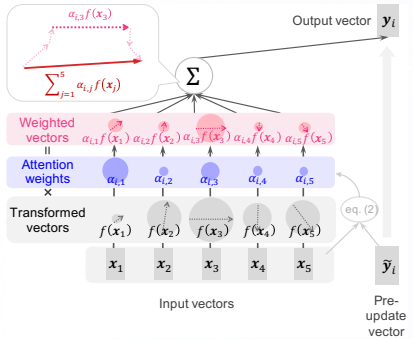
[ACL 2019, INLG 2019, EMNLP 2019, ACL 2020, Journal of CogSci 20, EMNLP-demo 2020, UIST 2023]

- Editor for natural language
- Startup



LM interpretability

[EMNMLP 2020, EMNLP 2021, ACL 2023, EMNLP 2023, ICLR 2024 (spotlight)]



What is language processing in humans? How can we help our own?

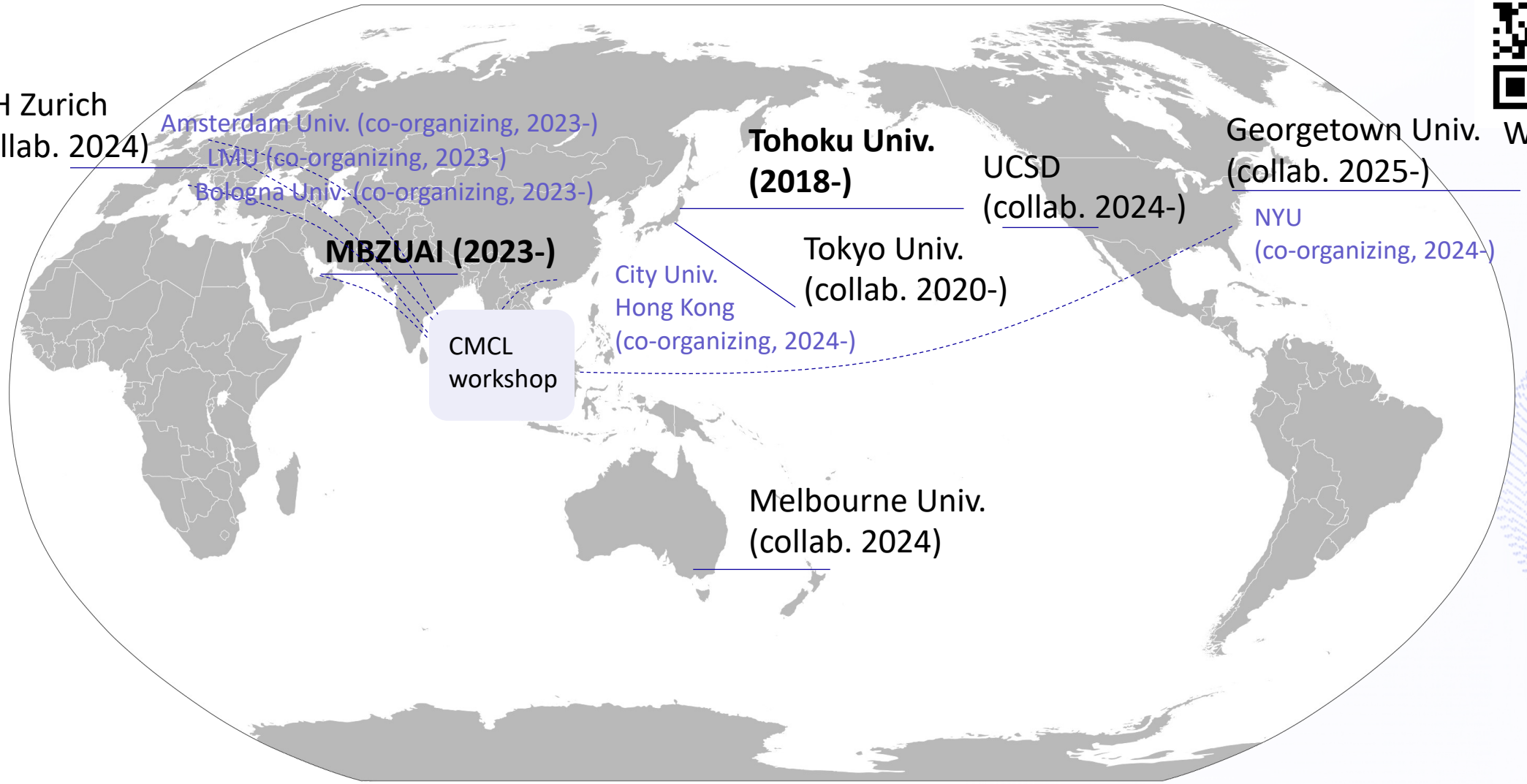
Computational psycholinguistics

[ACL 2021, EMNLP 2022, ACL 2023, COLING 2024, NAACL 2024, ACL 2024, COLING 2025]

(First-authored, co-authored)

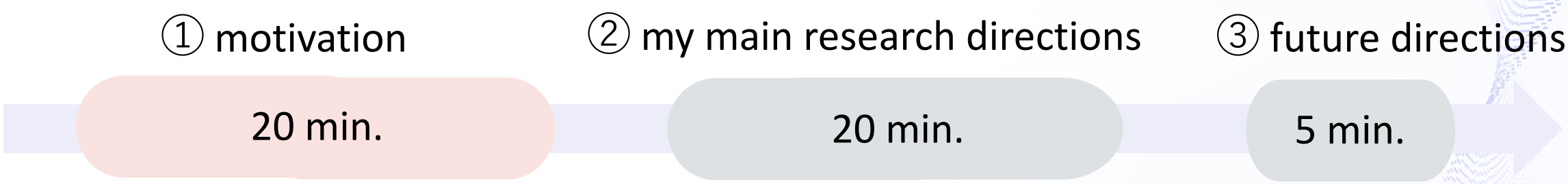
How different is language processing in humans and LMs?

International research connections



Web page

Roadmap


① motivation
② my main research directions
③ future directions
20 min.
20 min.
5 min.

Scientific modeling

- Why do planets move as observed?

- How did organic compounds emerge on the Earth?

- How do crowd crushes occur?

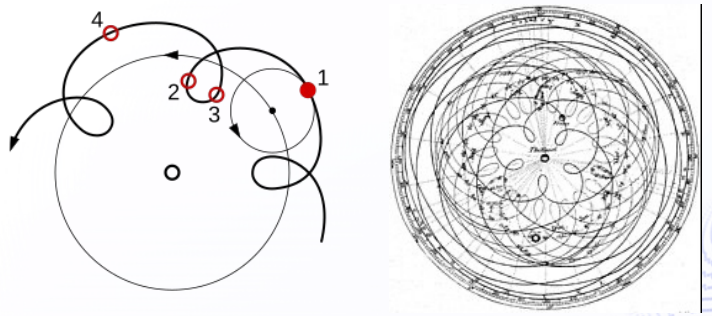
(figures are from Wikipedia or https://www.irasutoya.com/ , unless otherwise stated)

Scientific modeling

- Why do planets move as observed?

Build a model

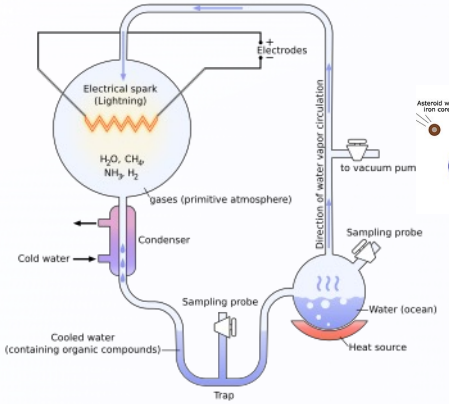
Test the fit/prediction

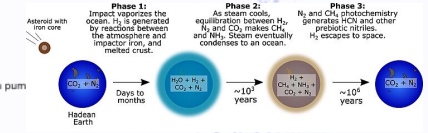


- How did organic compounds emerge on the Earth?

Build a model

Test the fit/prediction

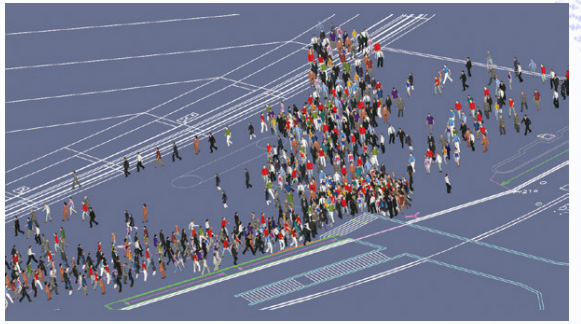




- How do crowd crushes occur?

Build a model

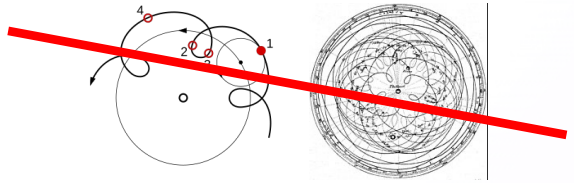
Test the fit/prediction



https://www.rcast.u-tokyo.ac.jp/ja/research/nishinari_lab.html

(figures are from Wikipedia or https://www.irasutoya.com/ , unless otherwise stated)

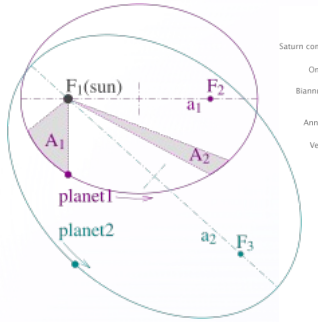
Scientific modeling

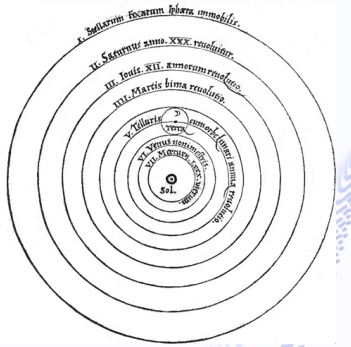


- Why do planets move as observed?

Build a model

Test the fit/prediction

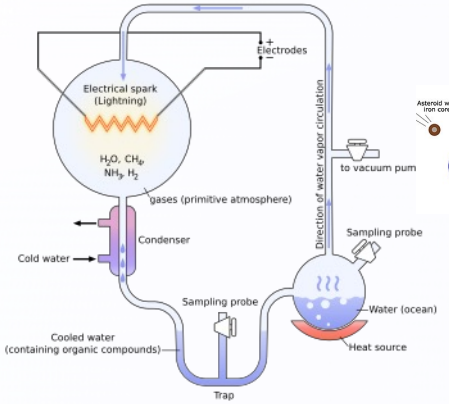


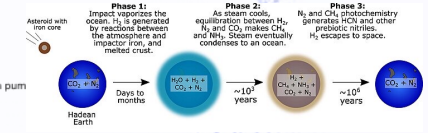


- How did organic compounds emerge on the Earth?

Build a model

Test the fit/prediction





- How do crowd crushes occur?

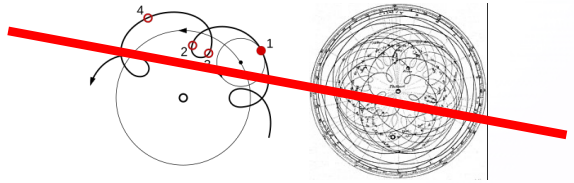
Build a model

Test the fit/prediction

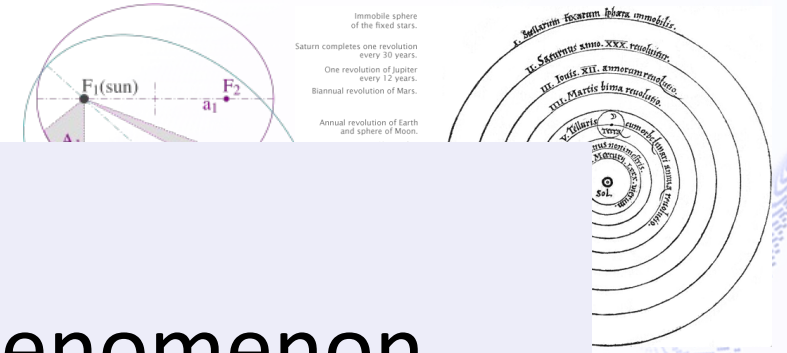
https://www.rcast.u-tokyo.ac.jp/ja/research/nishinari_lab.html

(figures are from Wikipedia or https://www.irasutoya.com/ , unless otherwise stated)

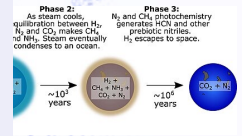
Scientific modeling



- Why do planets move as observed?



- How



- How

If a model exactly simulates a phenomenon of interest, the model serves as a good hypothesis/explanation for that phenomenon.

(aka. scientific modeling)

Test the fit/prediction

https://www.rcast.u-tokyo.ac.jp/ja/research/nishinari_lab.html

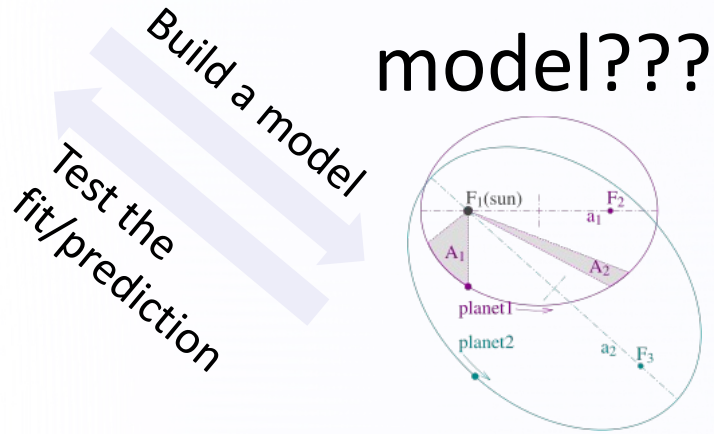
(figures are from Wikipedia or https://www.irasutoya.com/ , unless otherwise stated)

Fundamental linguistic questions

- What are humans computing during real-time language processing?
  - *What kind of equation are you now computing in front of this slide?*
- What is minimum requirements to be able to acquire language?
  - *Why do cats never start talking even if one keeps talking to them everyday?*
- Why do natural languages share certain universals, e.g., subject precedes objects?
  - *Why do languages shape as is? How did it emerge?*

Fundamental linguistic questions

- What are humans computing during real-time language processing?
  - *What kind of equation are you now computing in front of this slide?*
- What is minimum requirements to be able to acquire language?
  - *Why do cats never start talking even if one keeps talking to them everyday?*
- Why do natural languages share certain universals, e.g., subject precedes objects?
  - *Why do languages shape as is? How did it emerge?*



Why is artificial intelligence (AI) relevant to humanities?

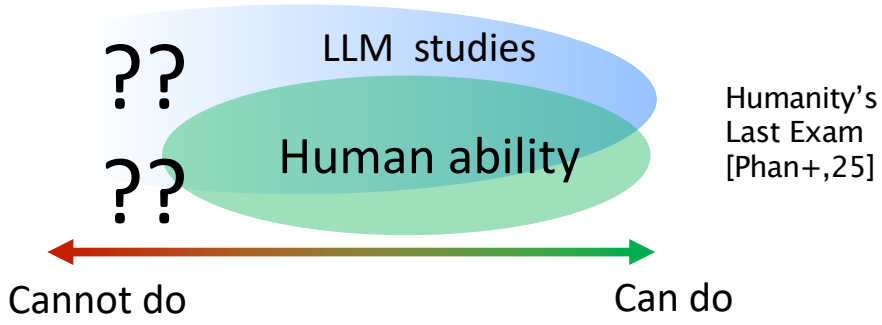
- Science requires objectivity
- Paradox: if humans start to introspect about ourselves to study human intelligence, this will lack objectivity
- Thus, we have to build a model (artificial intelligence), apart from humans and test it

- One of the original goals of the AI field --- understanding it by building it
  - *...the field (artificial intelligence) from three points of view: computational psychology, computational philosophy, and machine intelligence…The goal of computational psychology is* ***to understand human intelligent behavior by creating computer programs that behave in the same way that people do****… The program should do quickly what people do quickly, should do more slowly what people have difficulty doing, and should even tend to make mistakes where people tend to make mistake…* [Encyclopedia of Artificial Intelligence, Shapiro 1991]

Why is artificial intelligence (AI) relevant to humanities?

- Science requires objectivity
- Paradox: if humans start to introspect about ourselves to study human intelligence, this will lack objectivity
- Thus, we have to build a model (artificial intelligence), apart from humans and test it

- Here, the goal is to build an *exactly human-like* computational model that simulates phenomena of humans, following the scientific modeling approach



Humanity's Last Exam [Phan+,25]

Why is artificial intelligence (AI) relevant to humanities?

- Science requires objectivity
- Paradox: if humans start to introspect about ourselves to study human intelligence, this will lack objectivity
- Thus, we have to build a model (artificial intelligence), apart from humans and test it

Going back to 7 BCE - 16 CE…

- “Humans to explain humans” is super unethical (especially in causality experiments)

Pharaoh Psamtik
(664 – 610 BCE)

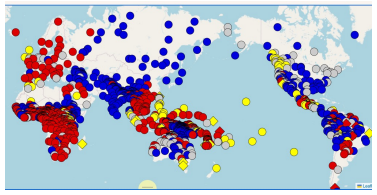
Frederick II
(1194-1250)

James IV
(1473-1513)

If one locks an infant in a room, what language will they start speaking?

(Thanks for Alex: https://gdr-lift.loria.fr/wp-content/uploads/2023/06/A.-Warstadt-ILFC-seminar-talk.pdf)

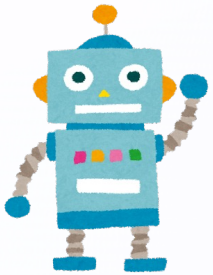
In 2025…



Human language

Build a model

Test the fit/prediction

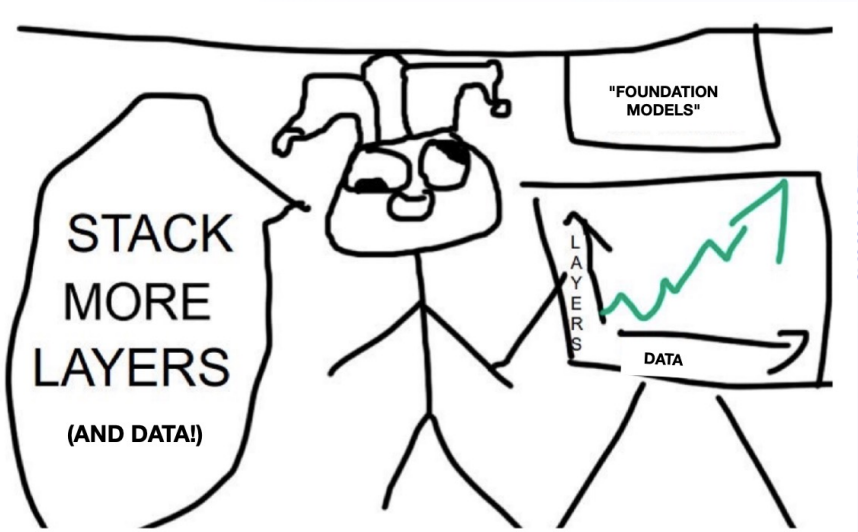


model

LLMs… are you the model of humans…?

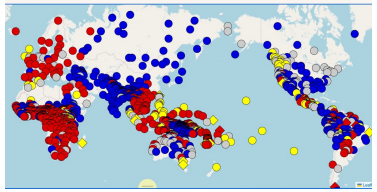
- We humans somehow found one way to build a model that behaves like humans

TL;DR

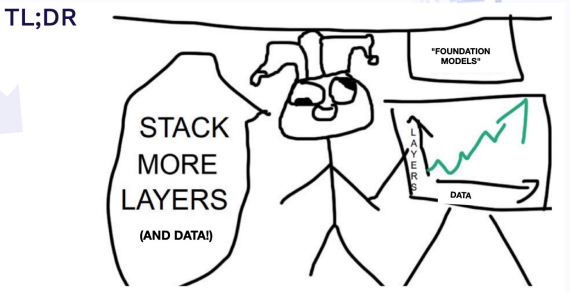


LLMs… are you the model of humans…?

- We humans somehow found one way to build a model that behaves like humans
- Some linguists criticize that this is not the model that linguistics has pursued

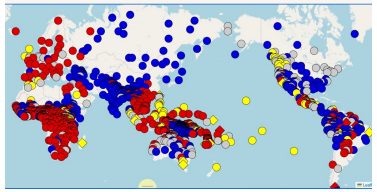


?????



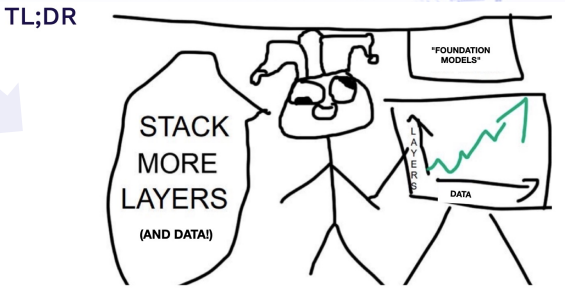
LLMs… are you the model of humans…?

- We humans somehow found one way to build a model that behaves like humans



- Some linguists criticize that this is not the model that linguistics has pursued

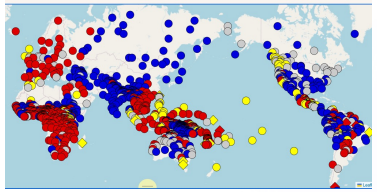
?????



- But we do not know other things that can learn human language as far as we know (in fact, it's seemingly working the best)
  - That's why NVIDIA stock is sparking

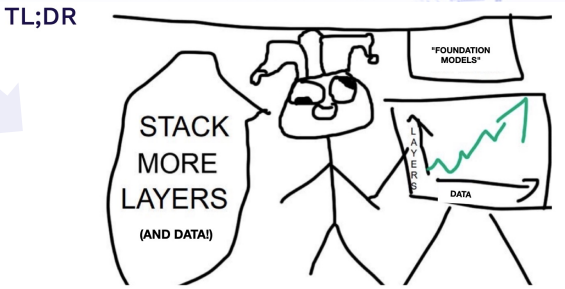
LLMs… are you the model of humans…?

- We humans somehow found one way to build a model that behaves like humans



- Some linguists criticize that this is not the model that linguistics has pursued

?????



- But we do not know other things that can learn human language as far as we know (in fact, it's seemingly working the best)
  - That's why NVIDIA stock is sparking

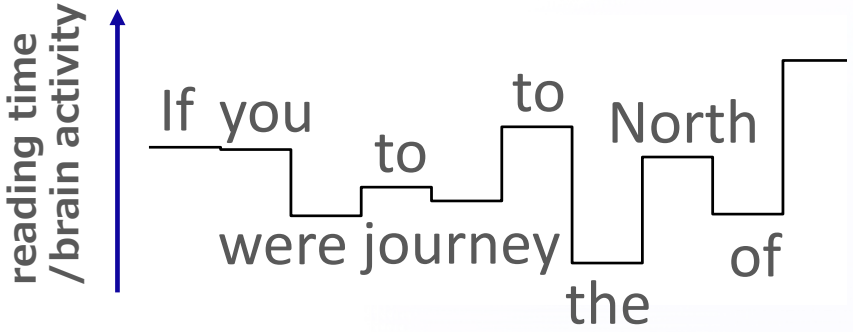
If a model exactly simulates a phenomenon of interest, the model serves as a good hypothesis/explanation for the phenomena.

(aka. scientific modeling)

Cognitive modeling

If you were to journey to the North of England, ⋯

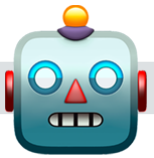
humans



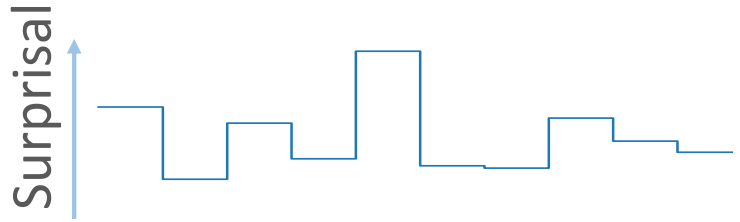
Tokens: $\boldsymbol{w} = \{w_1 \dots w_n\}$

Not tuning any part

If you were to journey to the North of England, ⋯



LM



Surprisal: $\hat{\boldsymbol{y}} = \{-\log_2 p(w_1|\boldsymbol{w}_{<1}) \dots -\log_2 p(w_n|\boldsymbol{w}_{<n})\}$

Unsupervised prediction*

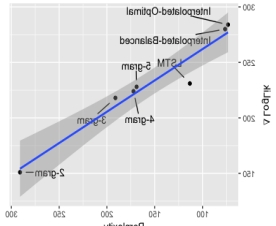
*training a regression model to rule out baseline factors and determine the coefficients, though

- The more unpredictable a word is, the more humans exhibit cognitive loads
  - The relationship should be logarithmic [Levy,08][Smith&Levy,13][Shain+,22]
  - Surprisal: $\text{Cost}(w_t) \propto -\log_2 p(w_t|\boldsymbol{w}_{<t-1})$

Are we approaching to the model of humans?
--- scaling law in cognitive modeling

LM-human correlations

[Frank&Bob, 2011]



[Goodkind&Bicknell, 2018]

???

Scaling up neural LMs

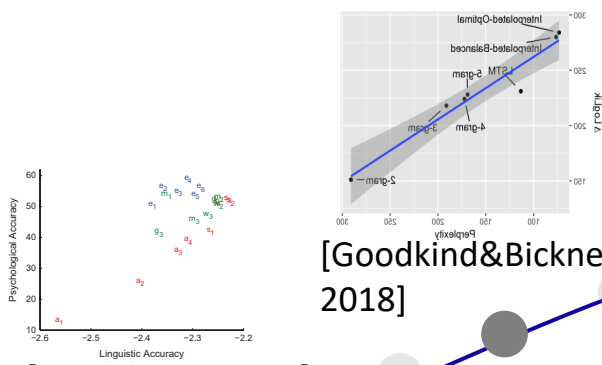
① motivation — 20 min.

② my main research directions — 20 min.

③ future directions — 5 min.

Are we approaching to the model of humans? --- scaling law in cognitive modeling

LM-human correlations



[Frank&Bob, 2011]

[Goodkind&Bicknell, 2018]

First pointed out this twist

Kuribayashi+21, ACL

Indeed decreasing!!

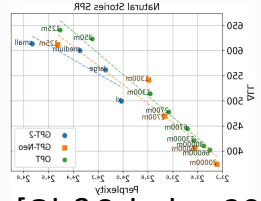
Kuribayashi+22, EMNLP

Why? 1

Yeah, why1 is main cause

Kuribayashi+24, NAACL

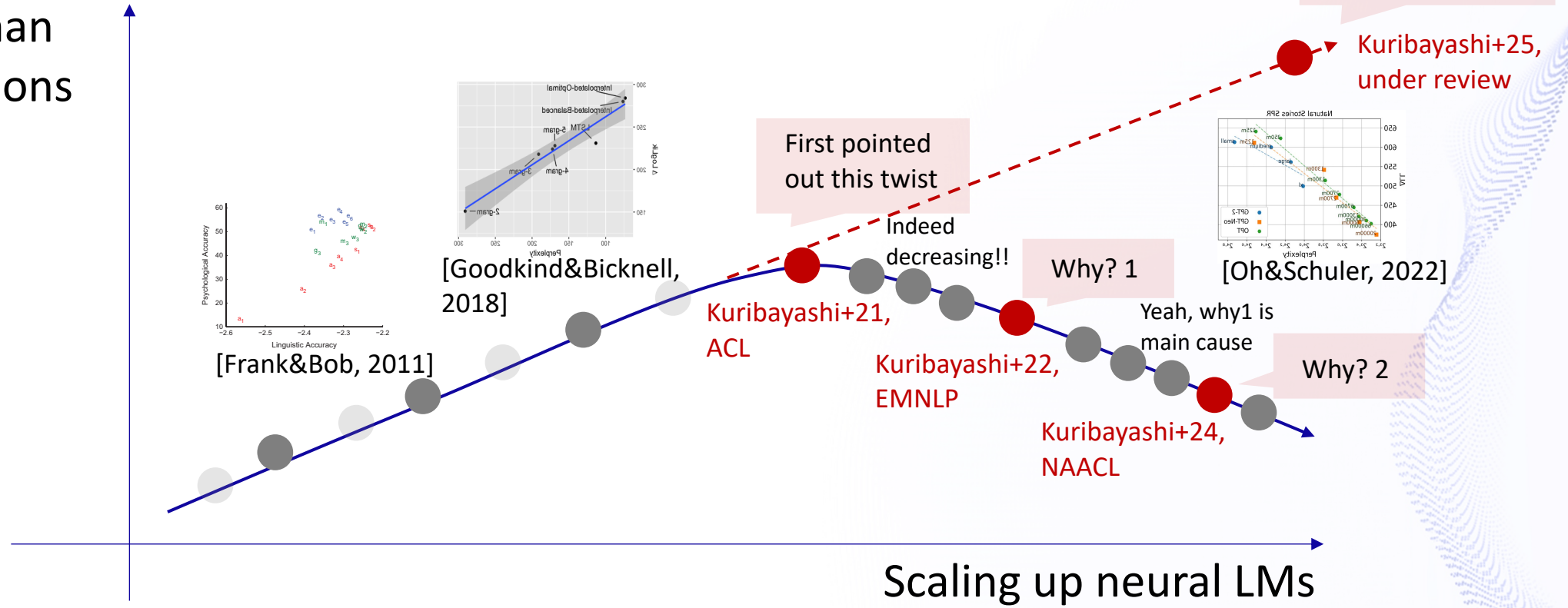
Why? 2



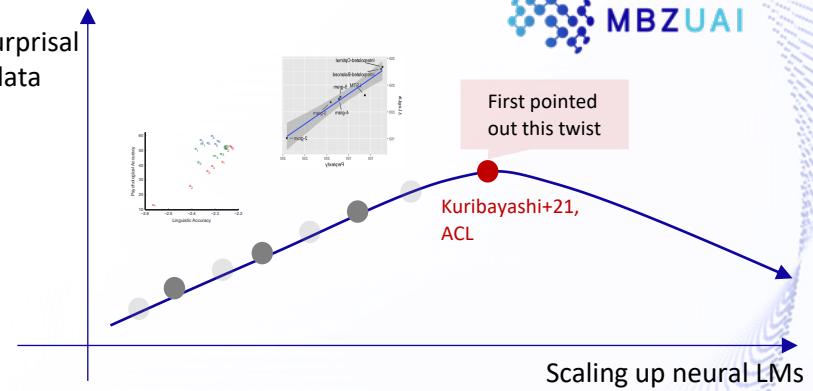
[Oh&Schuler, 2022]

Scaling up neural LMs

Are we approaching to the model of humans? --- scaling law in cognitive modeling



Kuribayashi+21 (ACL)

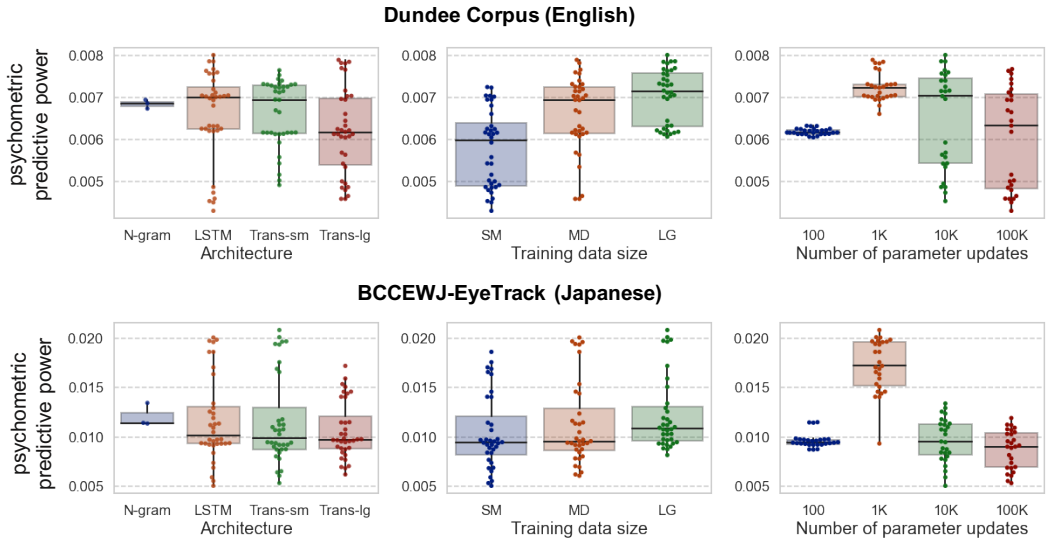


Lower Perplexity is Not Always Human-Like

Tatsuki Kuribayashi[1,2], **Yohei Oseki**[3,4], **Takumi Ito**[1,2],
Ryo Yoshida[3], **Masayuki Asahara**[5], **Kentaro Inui**[1,4]
[1]Tohoku University [2]Langsmith Inc. [3]University of Tokyo [4]RIKEN [5]NINJAL
{kuribayashi, takumi.ito.c4, inui}@tohoku.ac.jp ,
{oseki, yoshiryo0617}@g.ecc.u-tokyo.ac.jp ,masayu-a@ninjal.ac.jp

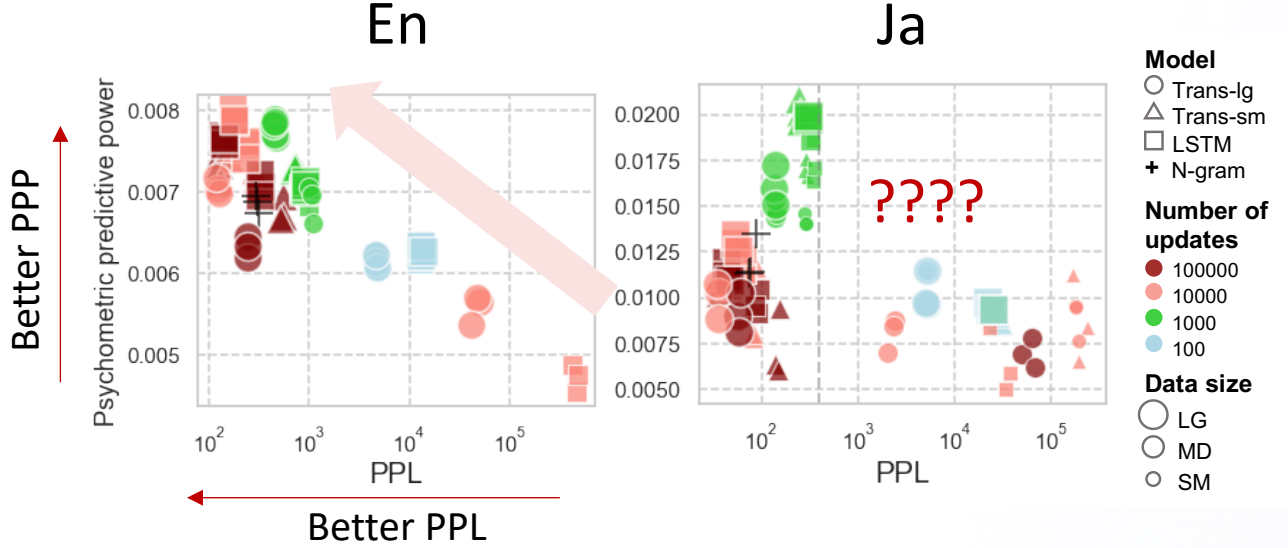
- First systematic, cross-linguistic evaluation of psychometric predictive power (PPP) of surprisal from neural LMs

$$\text{ReadingTime}(w_t) \propto -\log_{\theta} p(w_t|\boldsymbol{w}_{<t})$$

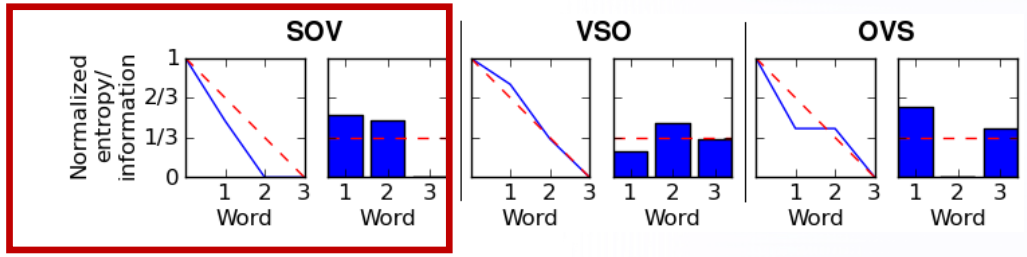


Kuribayashi+21 (ACL)

- **Previously reported monotonic relationship between LM scaling and PPP was fragile**



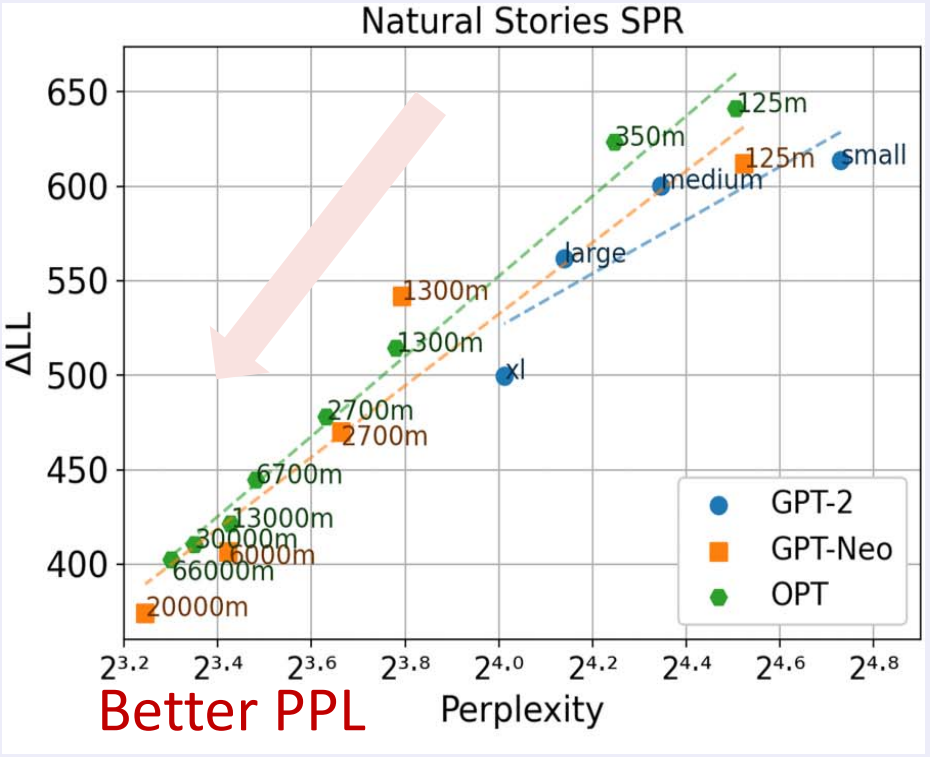
- Just changing the language (En->Ja) breaks it, empirically
  - Reading times and surprisals in the Japanese language (Subject-Object-Verb; SOV word order) have a large intra-sentential variance (i.e., low uniform information density), and LM-surprisal could not capture this variation well



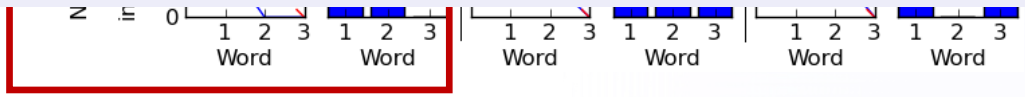
[Maurits+, 2010]

Kuribayashi+21 (ACL)

- With larger models, the negative scaling effect appeared even in the English language.
  - We could not observe it in our ACL 2021 work since we used in-house smaller LMs

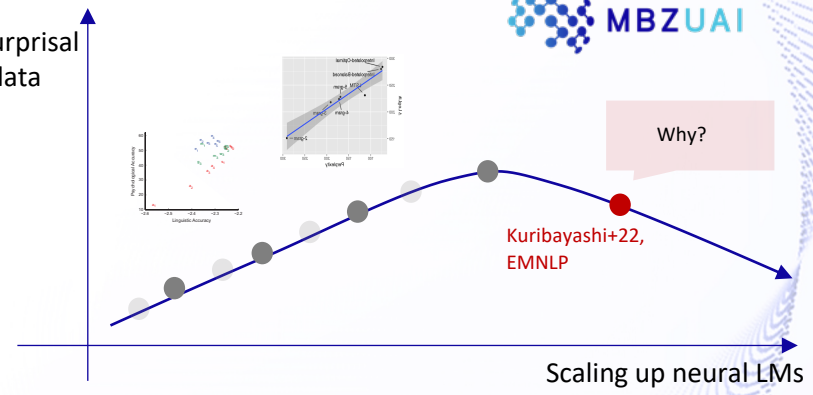


[Oh&Schuler, 2022]



[Maurits+, 2010]

Kuribayashi+22 (EMNLP)



Context Limitations Make Neural Language Models More Human-Like

Tatsuki Kuribayashi[1,2] **Yohei Oseki**[3,4] **Ana Brassard**[1,4] **Kentaro Inui**[1,4]
[1]Tohoku University [2]Langsmith Inc. [3]University of Tokyo [4]RIKEN
{kuribayashi, inui}@tohoku.ac.jp
oseki@g.ecc.u-tokyo.ac.jp ana.brassard@riken.jp

- Why did LMs' prediction deviate from humans?
- LMs (Transformers w/ self-attention) may be too good to consider wide contexts, compared to human real-time language processing

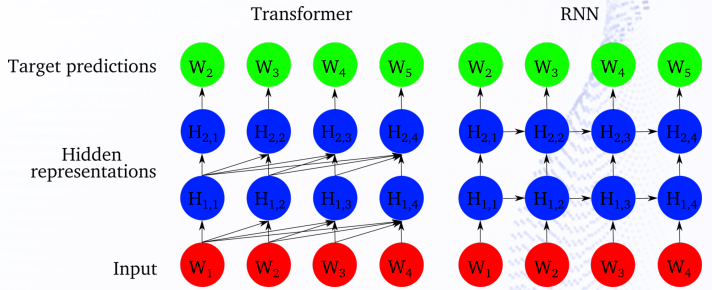


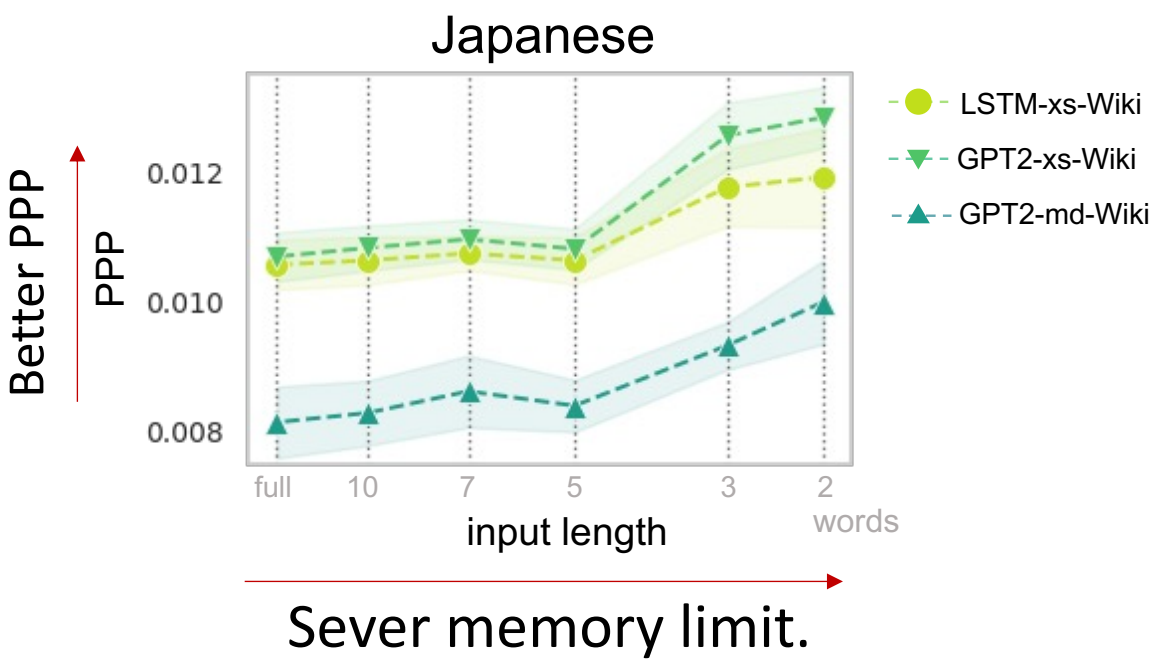
Figure 1: Comparison of sequential information flow through the Transformer and RNN, trained on next-word prediction.

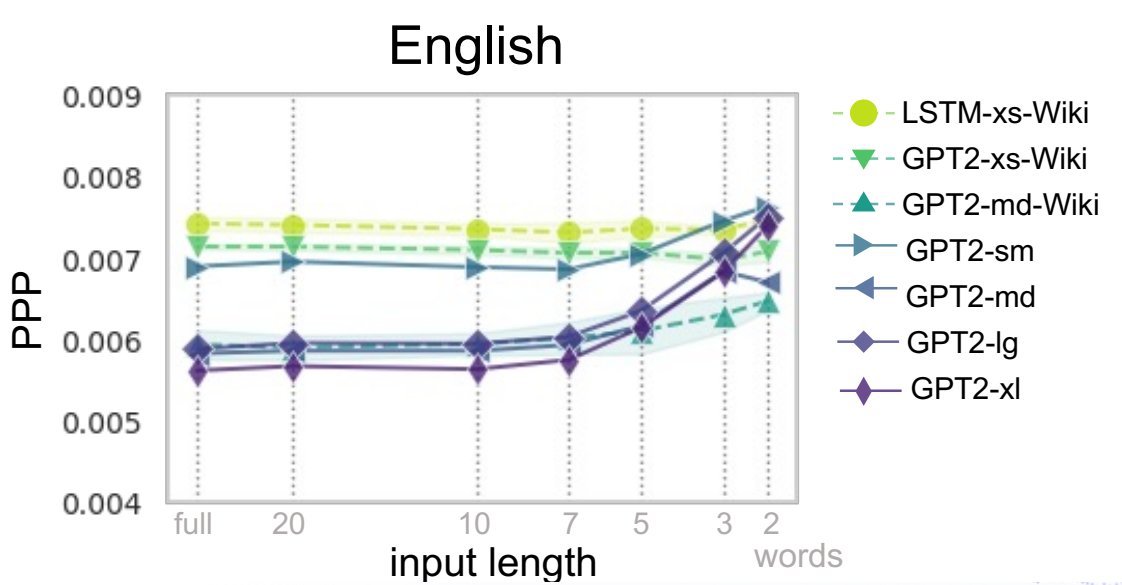
[Merkx&Frank, 21]

Kuribayashi+22 (EMNLP)

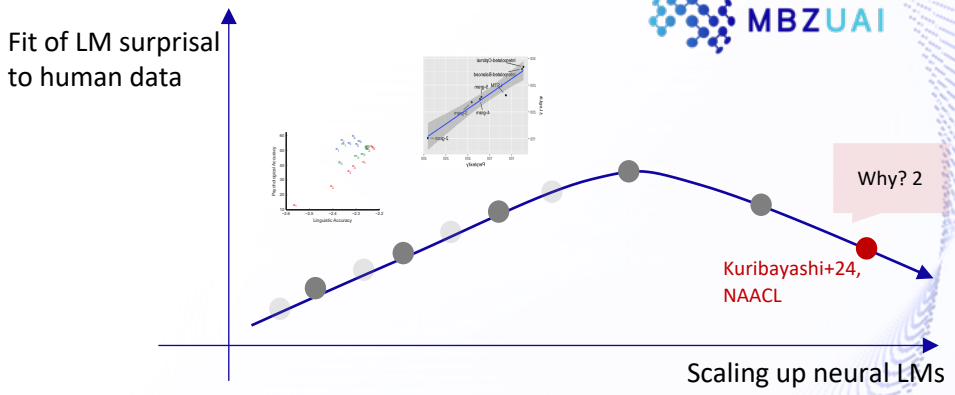
- Limiting LMs memory capacity aligns with humans
  - simple erasure of distant contexts surprisingly works well

$$\text{ReadingTime}(w_t) \propto -\cancel{-\log_2} p(w_t | w_0, w_1, \ldots w_{t-2}, w_{t-1})$$





Kuribayashi+24 (NAACL)

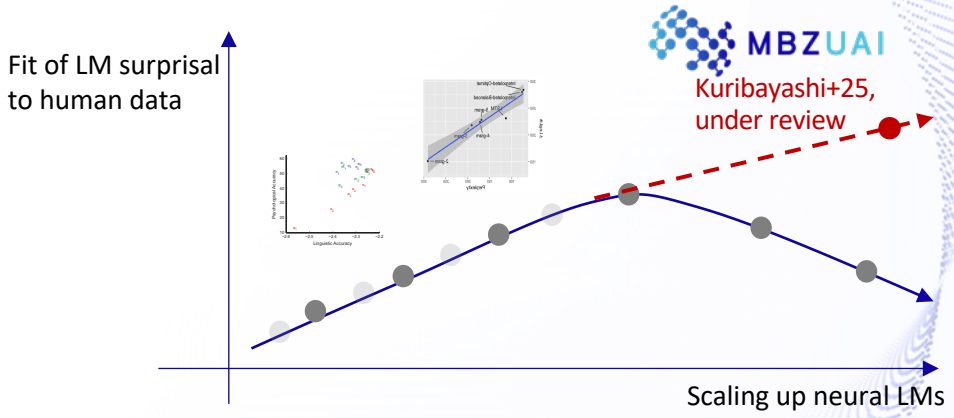


Psychometric Predictive Power of Large Language Models

Tatsuki Kuribayashi[1] **Yohei Oseki**[2] **Timothy Baldwin**[1,3]
[1]MBZUAI [2]The University of Tokyo [3]The University of Melbourne
{tatsuki.kuribayashi,timothy.baldwin}@mbzuai.ac.ae
oseki@g.ecc.u-tokyo.ac.jp

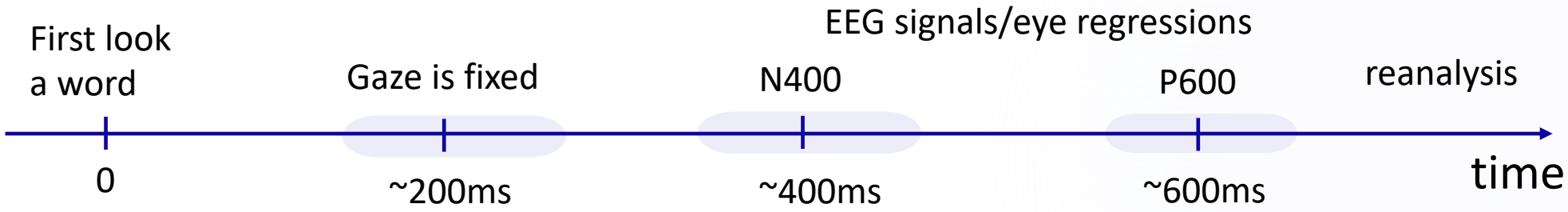
- Instruction-tuning and/or meta-linguistic prompting (*"Let's predict language processing cost!"*) did not improve PPP
- **Vanilla surprisal from base LMs (w/o tuning) predicts human data the best**
  - Human real-time processing seem to be simply tuned to statistics of next-word probability

Kuribayashi+25 (under review)



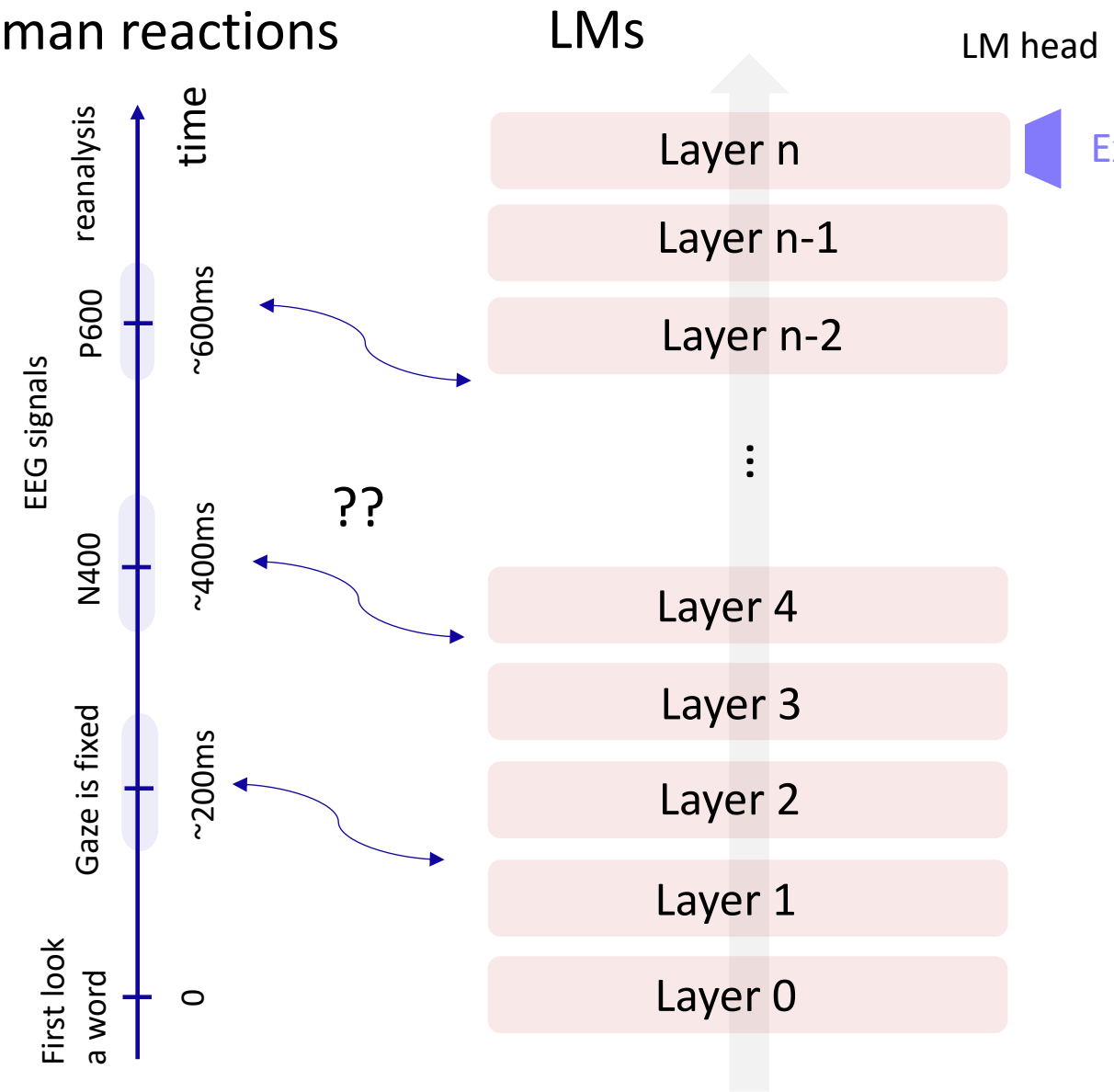
Large Language Models Are Human-Like Internally

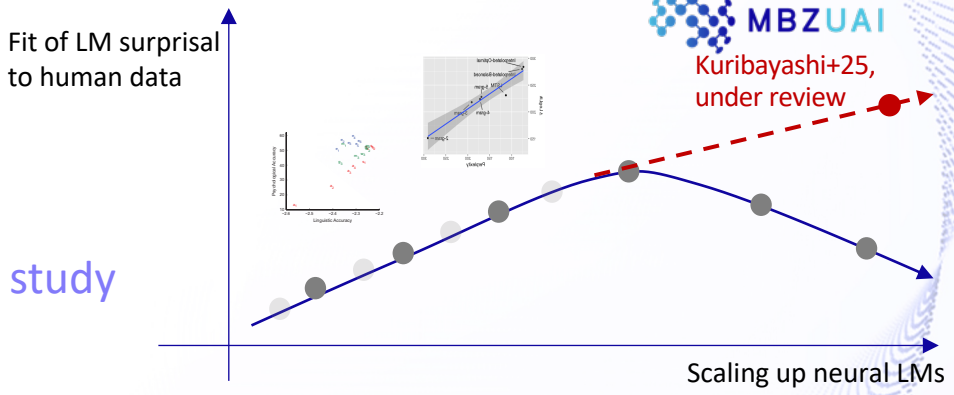
Tatsuki Kuribayashi[1] **Yohei Oseki**[2] **Souhaib Ben Taieb**[1,3]
Kentaro Inui[1,4,5] **Timothy Baldwin**[1,6]
[1]MBZUAI [2]The University of Tokyo [3]University of Mons
[4]Tohoku University [5]RIKEN [6]The University of Melbourne
{tatsuki.kuribayashi,souhaib.bentaieb,
kentaro.inui,timothy.baldwin}@mbzuai.ac.ae
oseki@g.ecc.u-tokyo.ac.jp



- Gaze duration is very fast (~200ms/word) and may reflect early-stage processing of language.
  - Where should such fast processing be realized in LLMs?

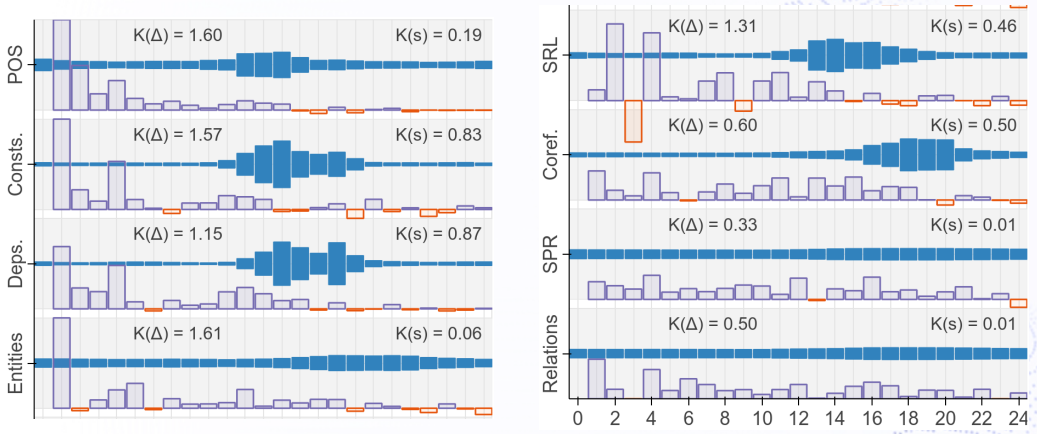
Kuribayashi+25 (under review)



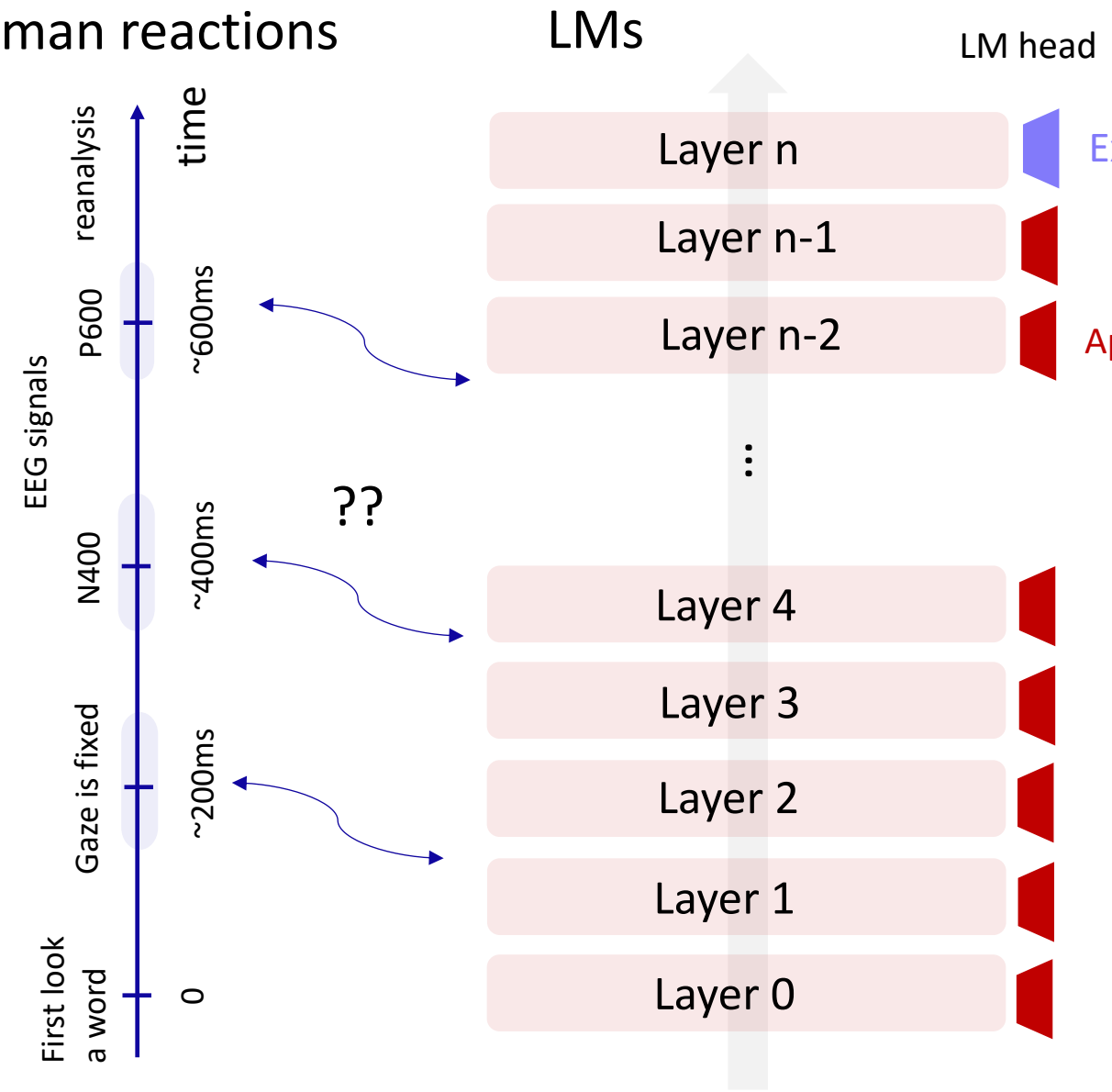


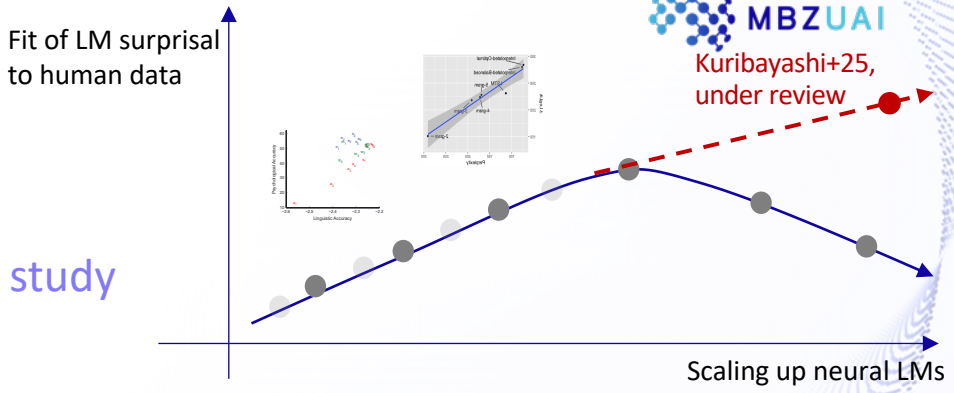
BERT Rediscovers the Classical NLP Pipeline

Ian Tenney[1] **Dipanjan Das**[1] **Ellie Pavlick**[1,2]
[1]Google Research [2]Brown University
{iftenney,dipanjand,epavlick}@google.com



Kuribayashi+25 (under review)

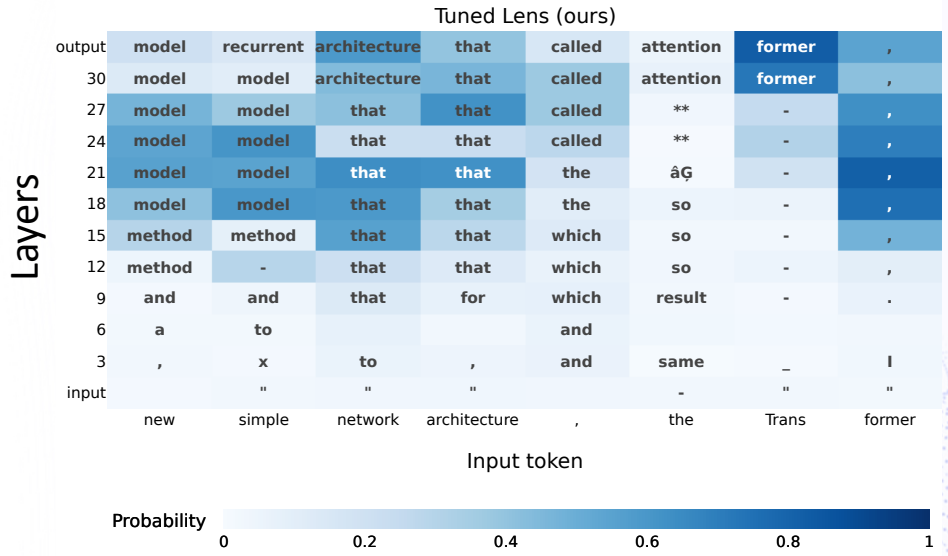




Existing study

Applying an LM-head to internal layers directly

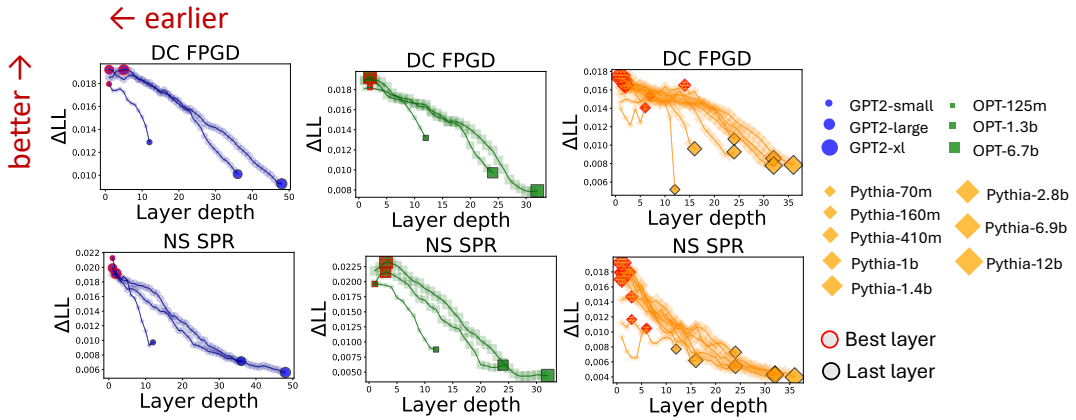
Surprisal $-\log p(w_t|\boldsymbol{w}_{<t}; \boldsymbol{h}_{t,l})$

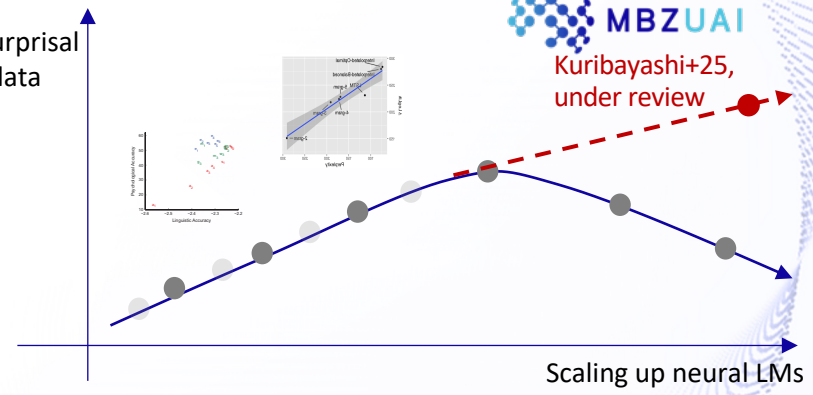


[Belrose+, 23]

Kuribayashi+25 (under review)

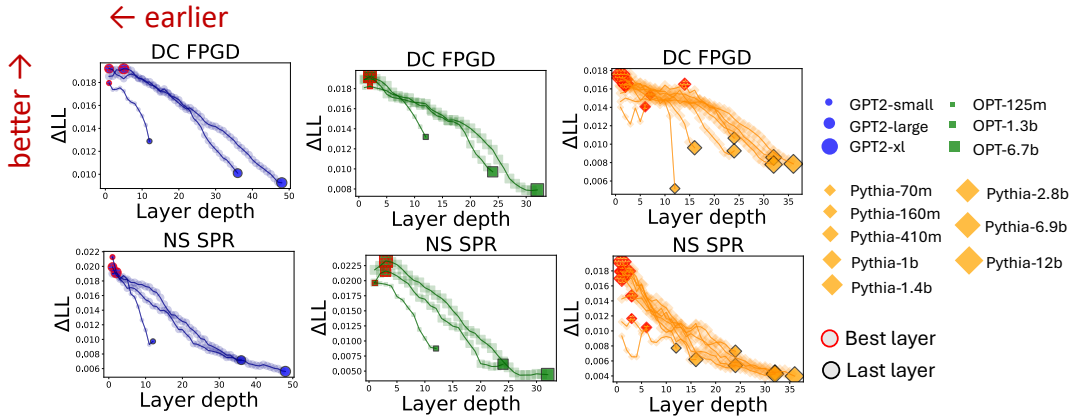
- (Fast) first pass gaze durations are better predicted in earlier layers



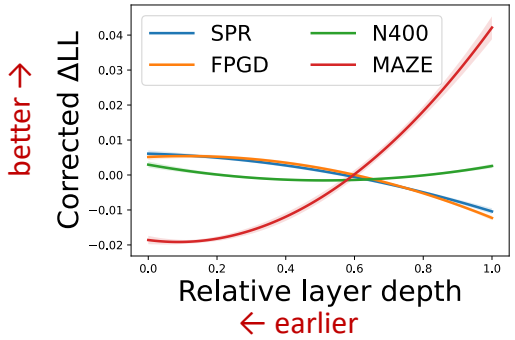


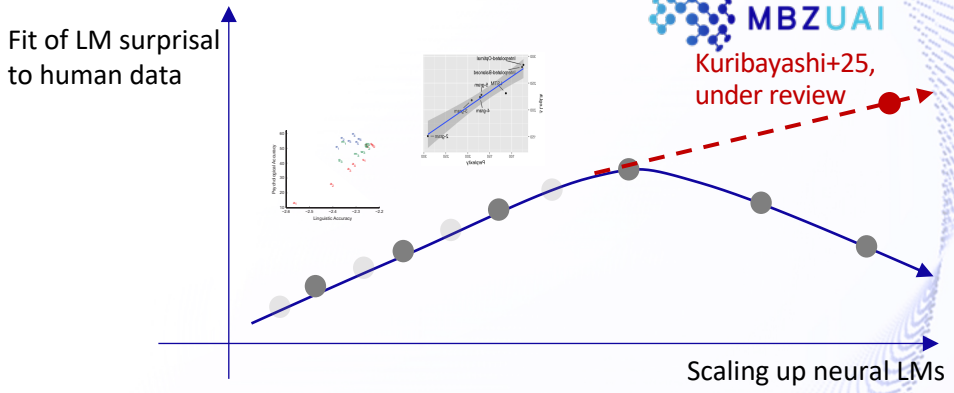
Kuribayashi+25 (under review)

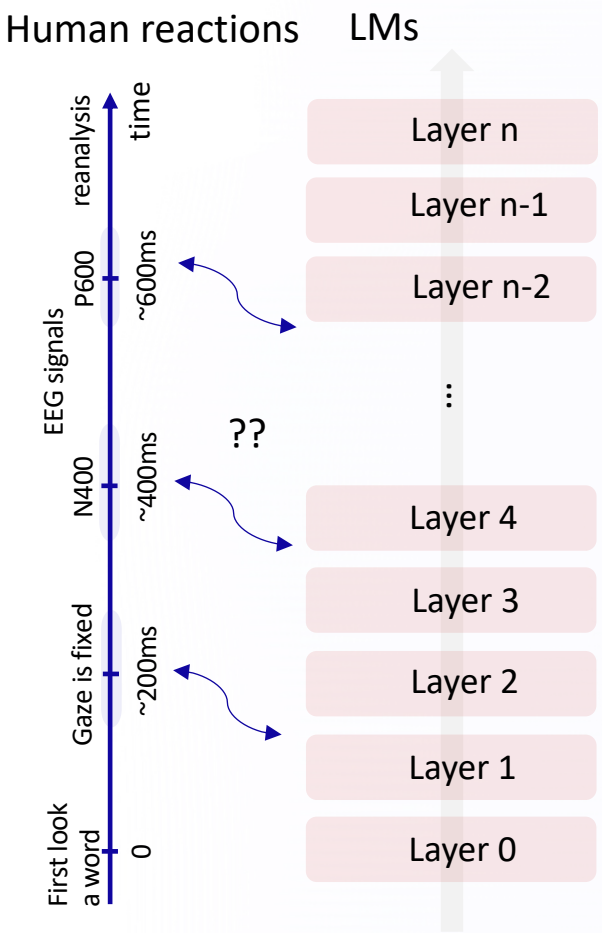
- (Fast) first pass gaze durations are better predicted in earlier layers



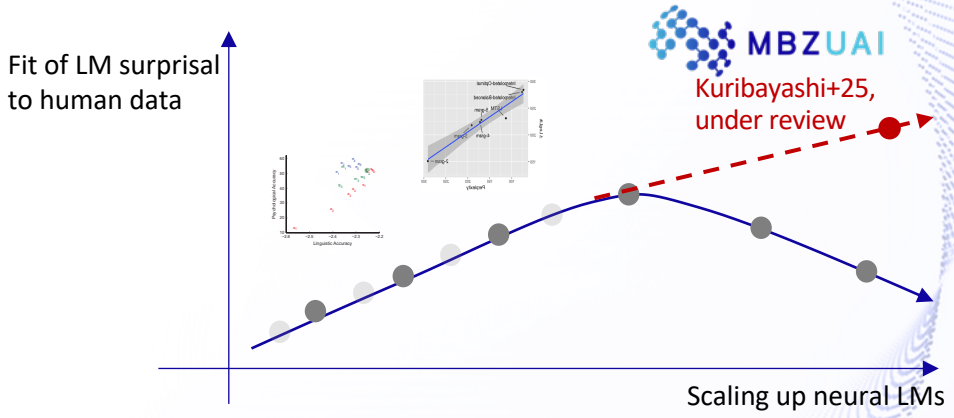
- Slower measures (N400, MAZE) tend to be better predicted in later layers



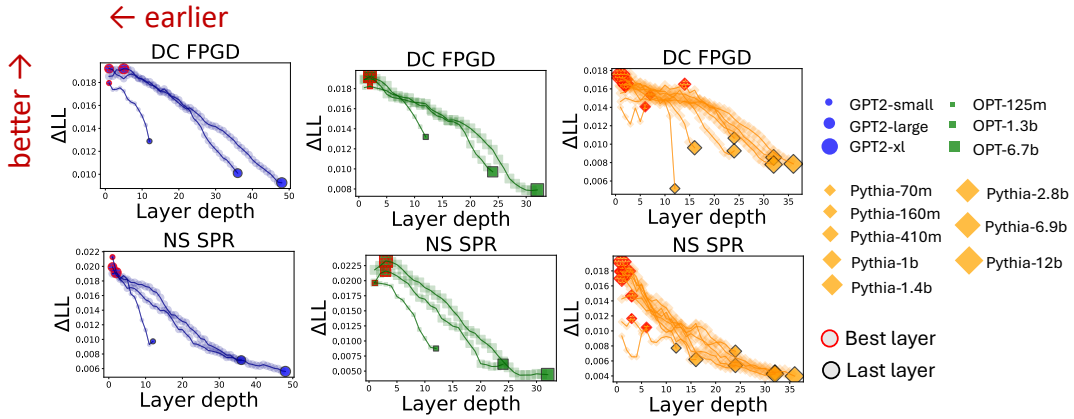




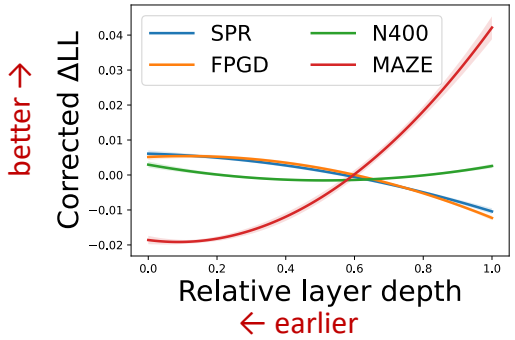
Kuribayashi+25 (under review)



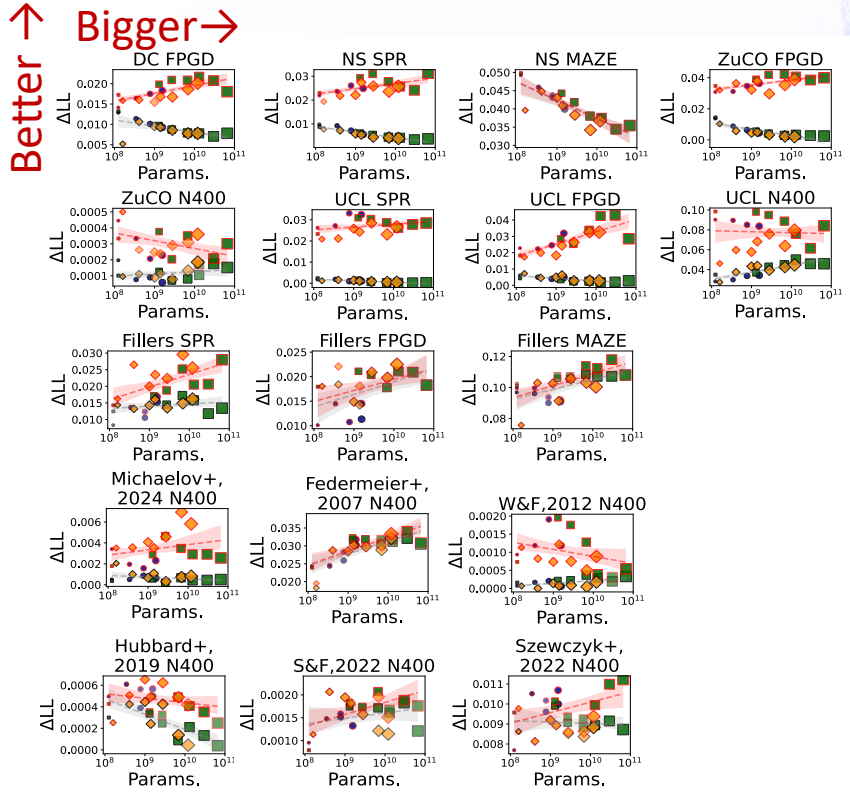
- (Fast) first pass gaze durations are better predicted in earlier layers



- Slower measures (N400, MAZE) tend to be better predicted in later layers



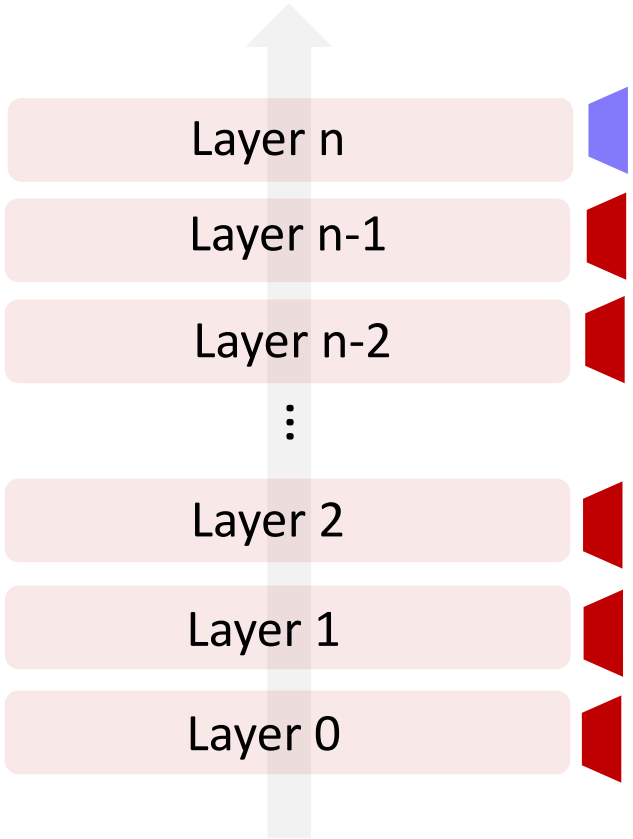
- Early layers in LLMs are human-like



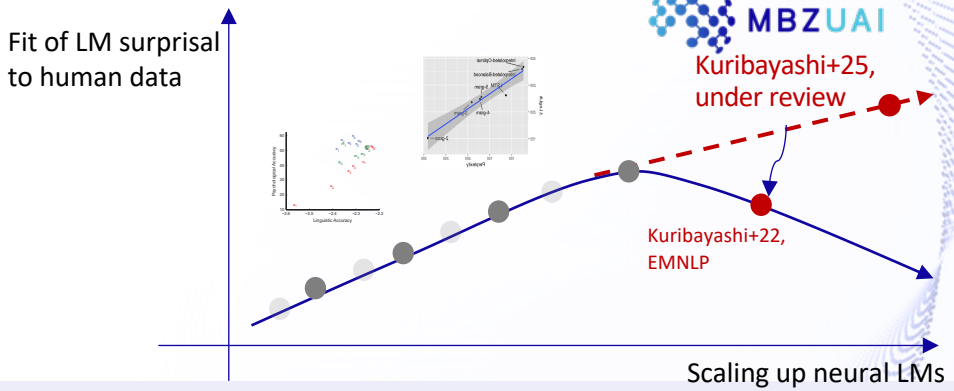
(To be continued…)

RE: Kuribayashi+22 (EMNLP)

- Moderately-contextualized, human-like surprisal from LLMs

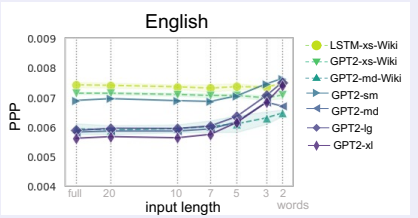


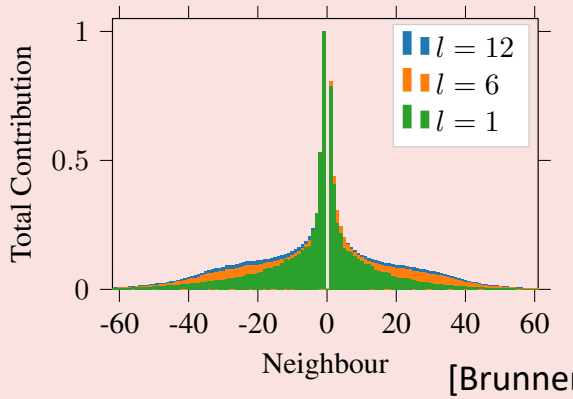
Human-like



Context Limitations Make Neural Language Models More Human-Like

Tatsuki Kuribayashi[1,2] **Yohei Oseki**[3,4] **Ana Brassard**[1,4] **Kentaro Inui**[1,4]
[1]Tohoku University [2]Langsmith Inc. [3]University of Tokyo [4]RIKEN
{kuribayashi, inui}@tohoku.ac.jp
oseki@g.ecc.u-tokyo.ac.jp ana.brassard@riken.jp





[Brunner+, 19]

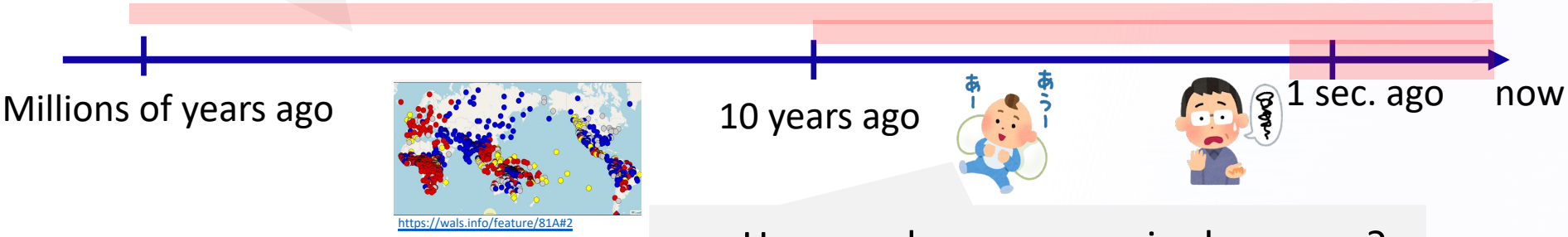
Earlier layers are less-contextualized

Fundamental linguistic problems

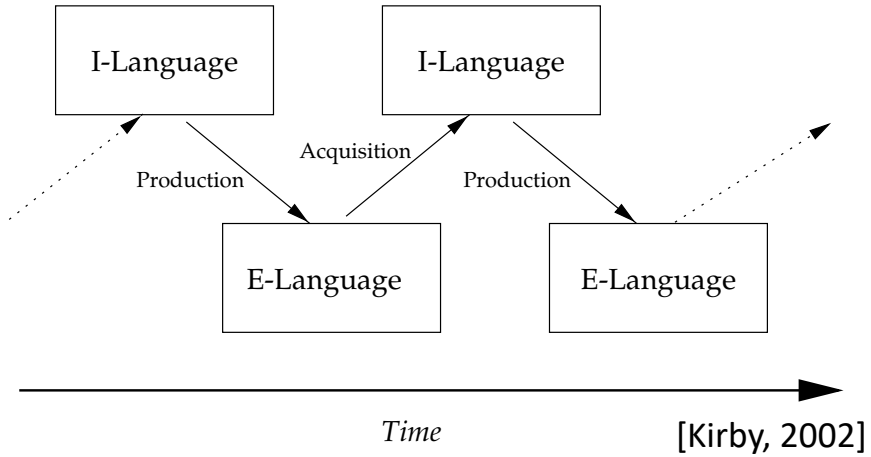
- Why do natural languages have typological universals, e.g., subject≺object?

Language easy to process would have survived

- What are humans computing during real-time language processing?



- How can humans acquire language?

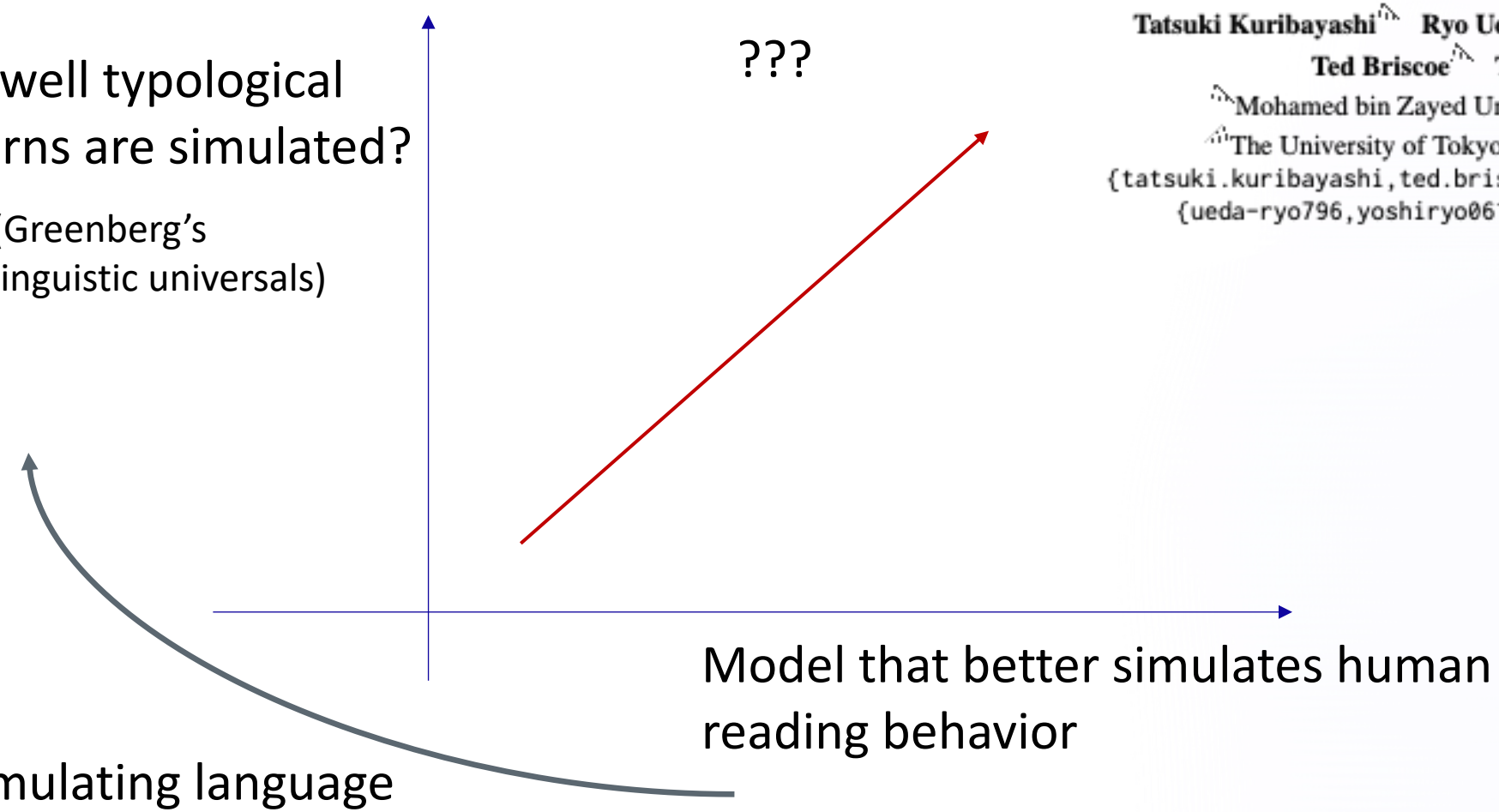


[Kirby, 2002]

From cognitive modeling to language universals Kuribayashi+24 (ACL)

How well typological patterns are simulated?

(Greenberg's linguistic universals)



Emergent Word Order Universals from Cognitively-Motivated Language Models

Tatsuki Kuribayashi[♠♡] Ryo Ueda[♣] Ryo Yoshida[♣] Yohei Oseki[♣]
Ted Briscoe[♡] Timothy Baldwin[♠♢]
[♠]Mohamed bin Zayed University of Artificial Intelligence
[♣]The University of Tokyo [♢]The University of Melbourne
{tatsuki.kuribayashi,ted.briscoe,timothy.baldwin}@mbzuai.ac.ae
{ueda-ryo796,yoshiryo0617,oseki}@g.ecc.u-tokyo.ac.jp

From cognitive modeling to language universals Kuribayashi+24 (ACL)

- A problem to predict the *plausibility* of language design, based on their learnability and processing difficulty for LMs.

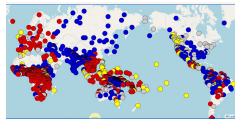
Toy languages generated by CFG

[subject precedes object, subject precedes verb, Adjective precedes noun…]

[subject follows object, subject precedes verb, Adjective precedes noun…]

…

[subject follows object, subject follows verb, Adjective follows noun…]



From cognitive modeling to language universals Kuribayashi+24 (ACL)

- A problem to predict the *plausibility* of language design, based on their learnability and processing difficulty for LMs.

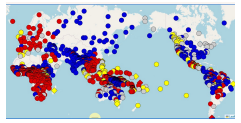
Toy languages generated by CFG

[subject precedes object, subject precedes verb, Adjective precedes noun…]

[subject follows object, subject precedes verb, Adjective precedes noun…]

…

[subject follows object, subject follows verb, Adjective follows noun…]



compare

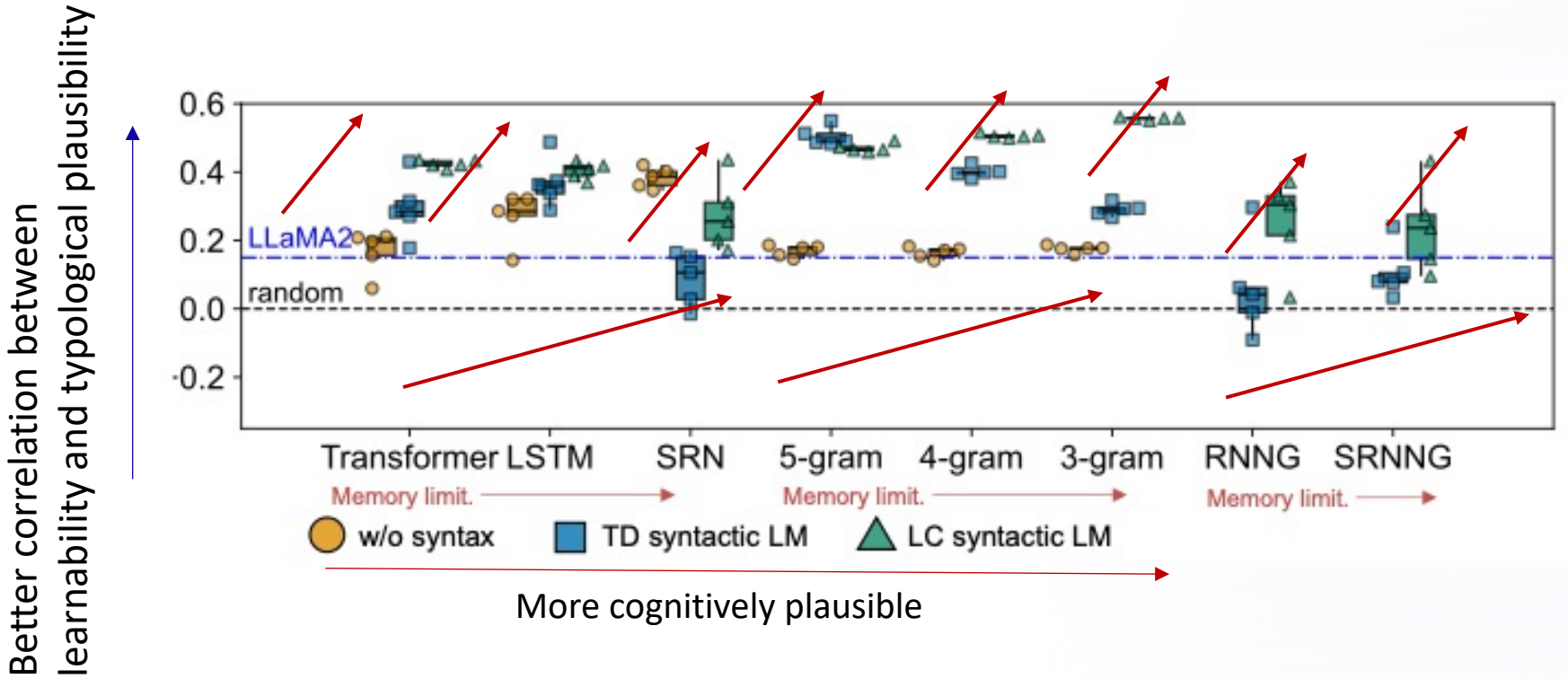
- Train LMs in each language and obtain learnability distribution across language configurations

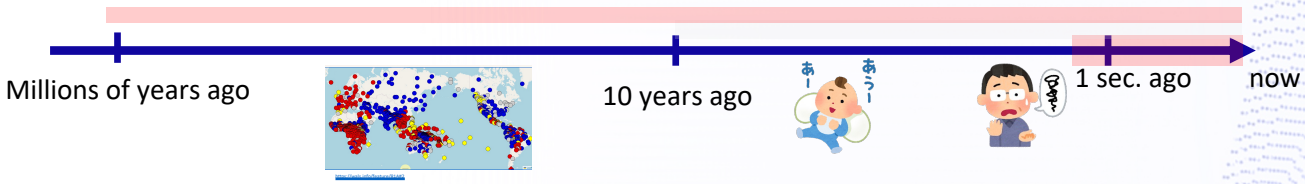
For NLPers: negative PPL; higher is better

- Which language is easier to learn for particular LMs?
  - Human-like LMs: memory limitation, syntax-aware, cognitively-plausible left-corner traversals

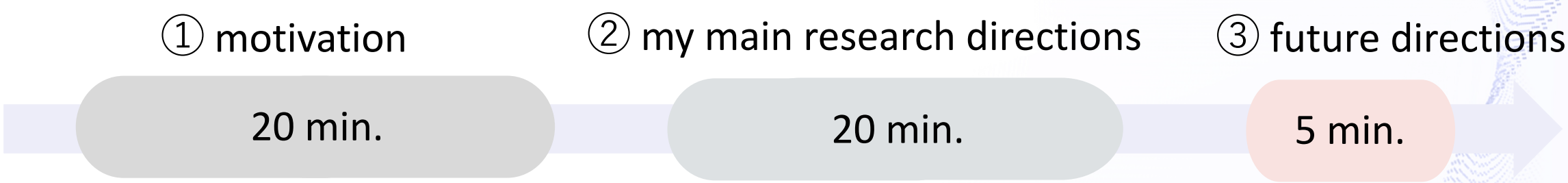
From cognitive modeling to language universals Kuribayashi+24 (ACL)

- Learning/processing difficulties of LMs are better correlated with typological distributions when using more cognitively-motivated LMs





Roadmap


① motivation
② my main research directions
③ future directions
20 min.
20 min.
5 min.

Future: Emergent corpus

- LLMs are good at learning language, **if there is a corpus**
  - Language transmission in one generation

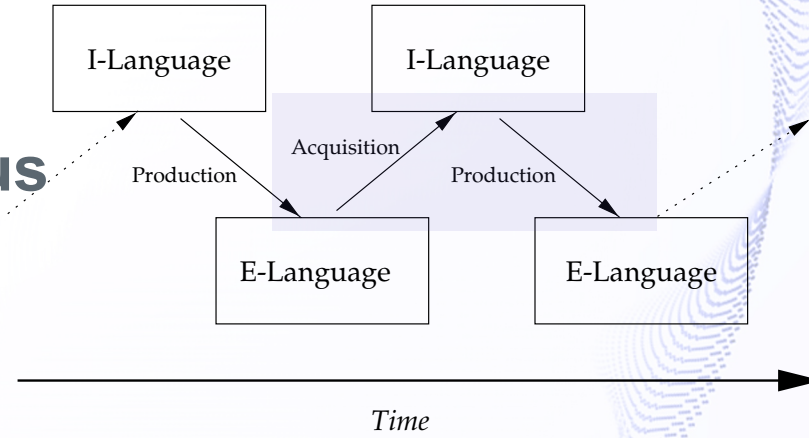
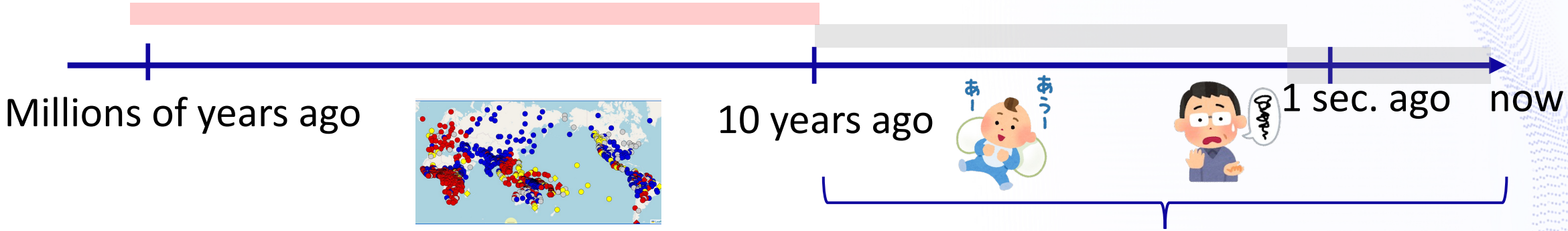


figure 6.1. The transmission of language over time.
[Kirby, 2002]

- Then, where is corpus from?
  - Humans have achieved LLM-like behaviors from a situation w/o corpus, in the long history on Earth
  - Connection to emergent language/communication/symbols
    - must be handled via computational simulation (computational linguistics!)



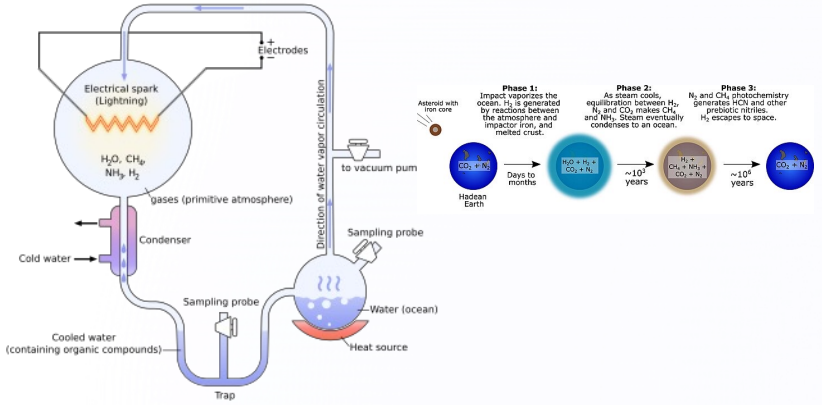
LLM training may give us a hint

Future: Connections to Robotics

- Text-only NLP alone can not explore language emergence under text-less environments
  - Agents should play real, physical games to explore the emergence of language

Build a model

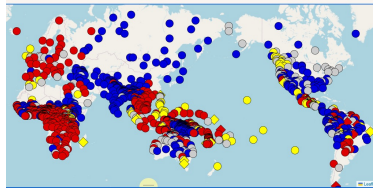
Test the fit/prediction



- If we can train LMs (robots) under the same scenarios as humans, and if they acquire language in the same way as humans, what does this imply?

Future: How should we measure human-likeness of LLMs?

- Humanities studies as checklists



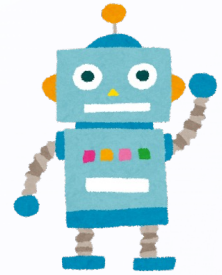
Human language

Describe/hints

Humanities studies

Build a model

Test the fit/prediction

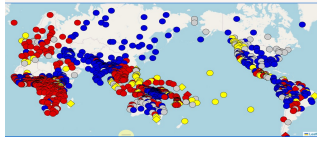


model

- What is minimum criteria to explain empirical linguistic observations?

- Can neural agents re-invent human language?
- Can LMs mimic human-like language acquisition patterns?
- Can LMs simulate human real-time language processing?
- Do LLMs have linguistic knowledge?

Millions of years ago



10 years ago

1 sec. ago

now

Future: Maintaining the community

- ~90% of (young) NLP researchers may be thinking about LLMs and chatbot
  - It may be rational, considering the current trend/economy, instead of exploring niche topics
  - How can community think more freely about diverse things or how can I encourage such activities? (I also sometimes feel a sense of isolation in the community)
    - The microwave oven was invented thanks to a person who happened to notice a melted chocolate in radar research.

- How to appeal the excitement of exploring scientific (humanities) questions?
- Isn't it only natural that we want to know about humans because we are humans?

- AI is not only for the science of artificial intelligence but also for any science using artificial intelligence