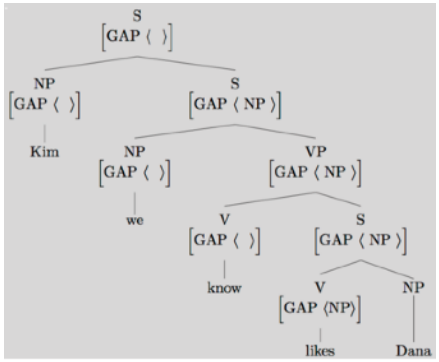
The Acquisition and Processing of Grammatical Structure: Insights from Deep Learning

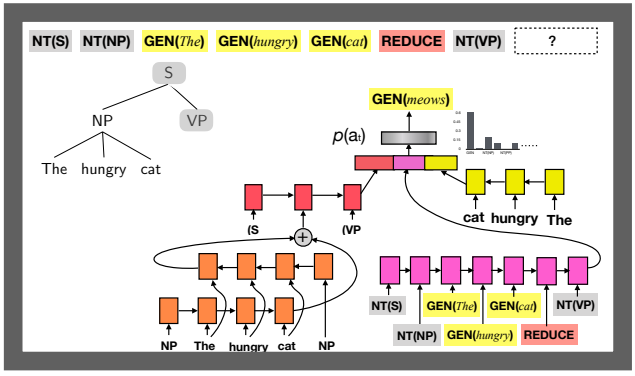
Roger Levy
Computational Psycholinguistics Laboratory (CPL)
Dept. of Brain & Cognitive Sciences
Massachusetts Institute of Technology

ILFC
Université de Paris
17 May 2022

Triangulating on a model of human(-like) language



Theory of linguistic knowledge

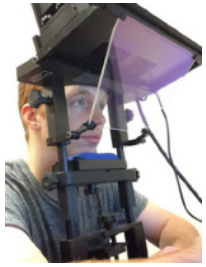


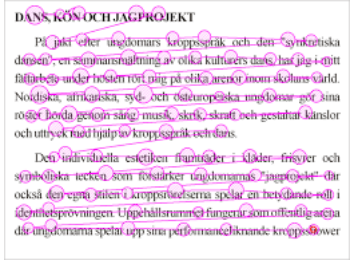
Computational Models

Human(-like) linguistic knowledge and use

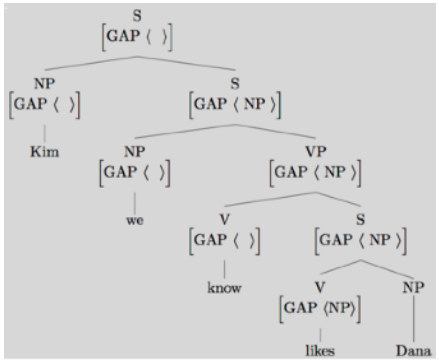
Language Datasets

Psychological Experimentation

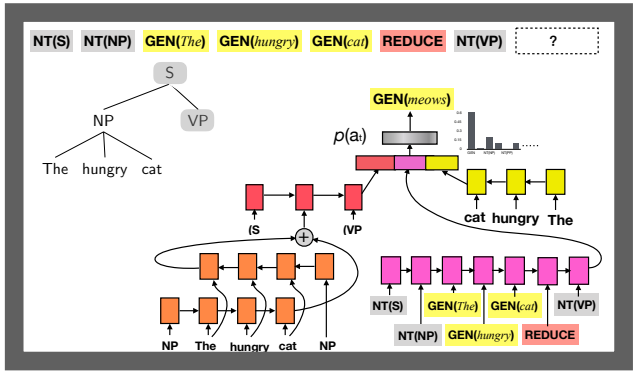




Triangulating on a model of human(-like) language



Theory of linguistic knowledge

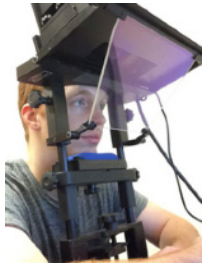


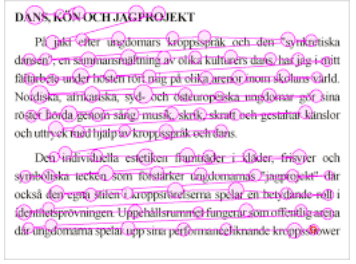
Computational Models

Human(-like) linguistic knowledge and use

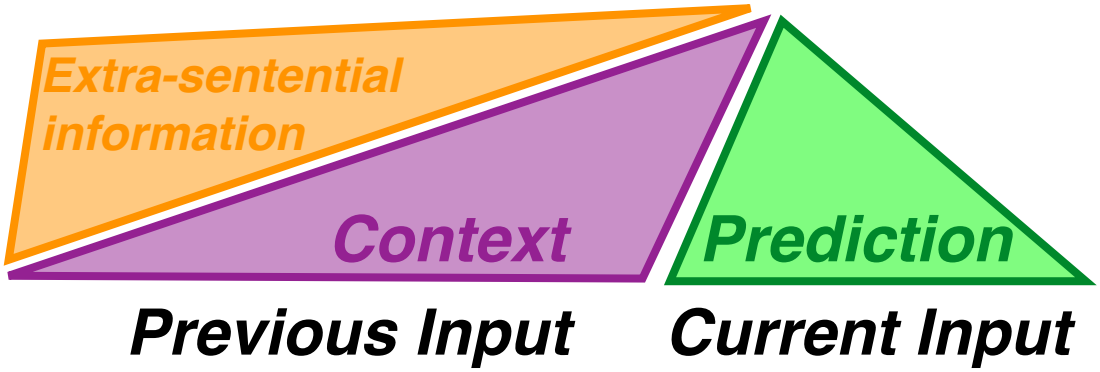
Language Datasets

Psychological Experimentation

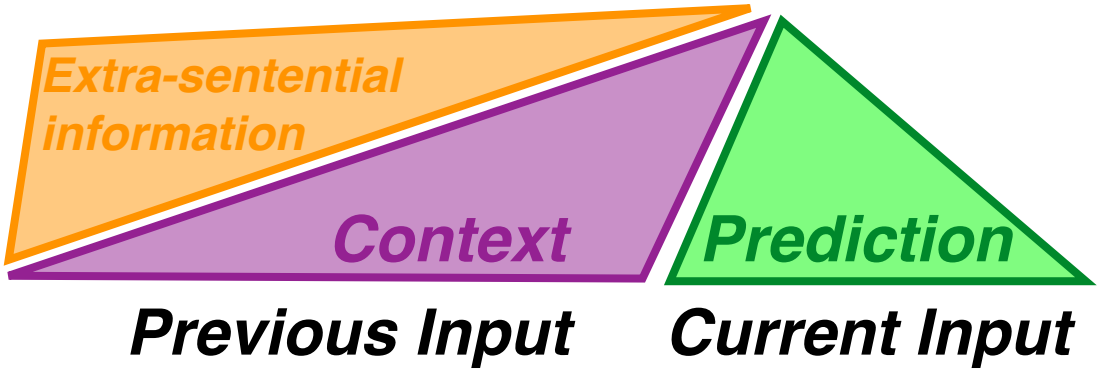




Expectations in incremental comprehension



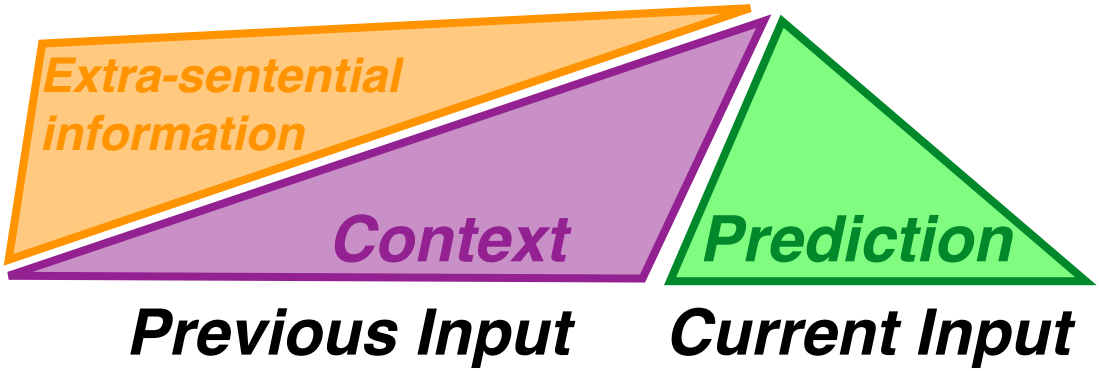
Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated…

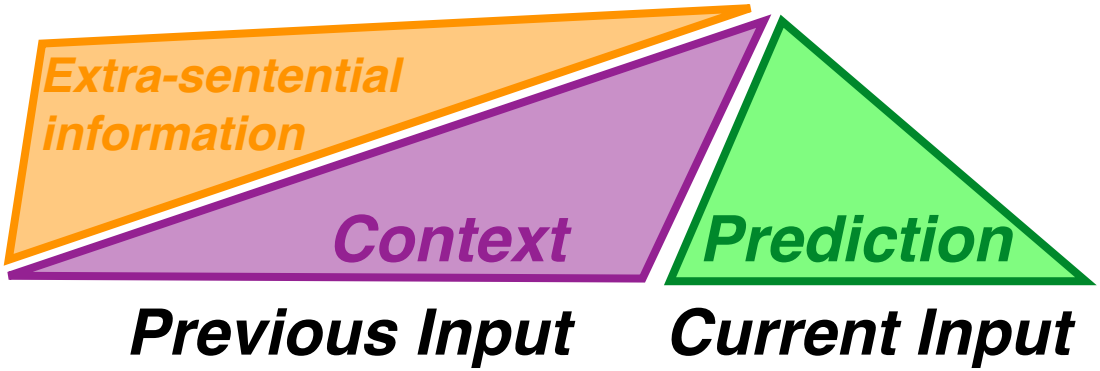
Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

Expectations in incremental comprehension



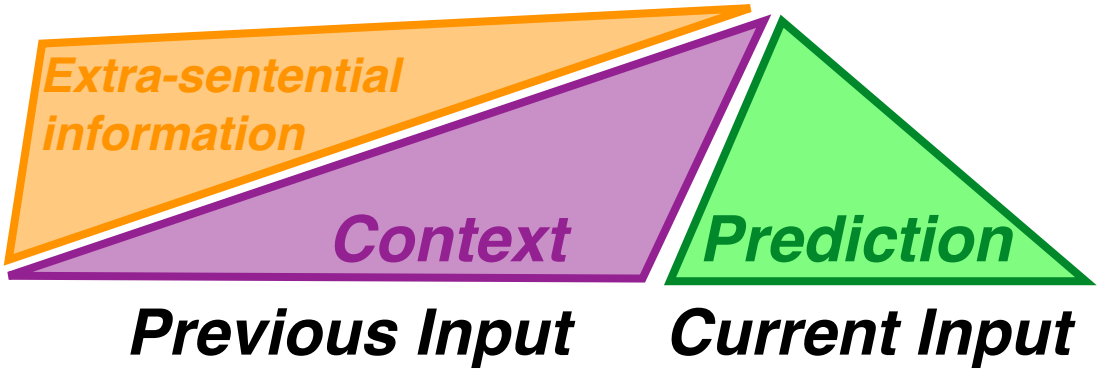
- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an…

Expectations in incremental comprehension



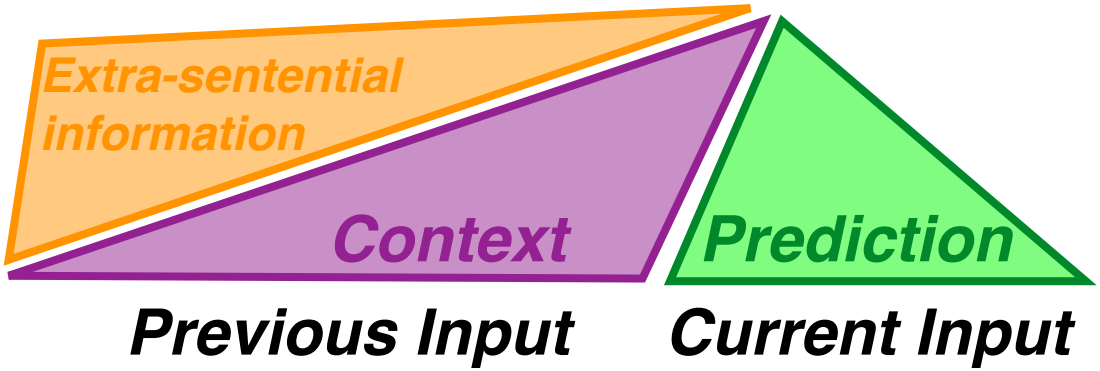
- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

Expectations in incremental comprehension



- Syntactic:

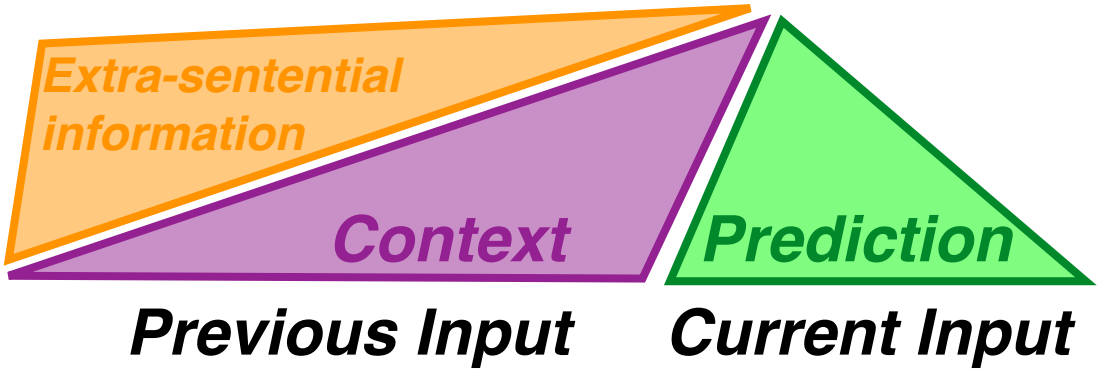
  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

  Terry ate a…

Expectations in incremental comprehension



- Syntactic:

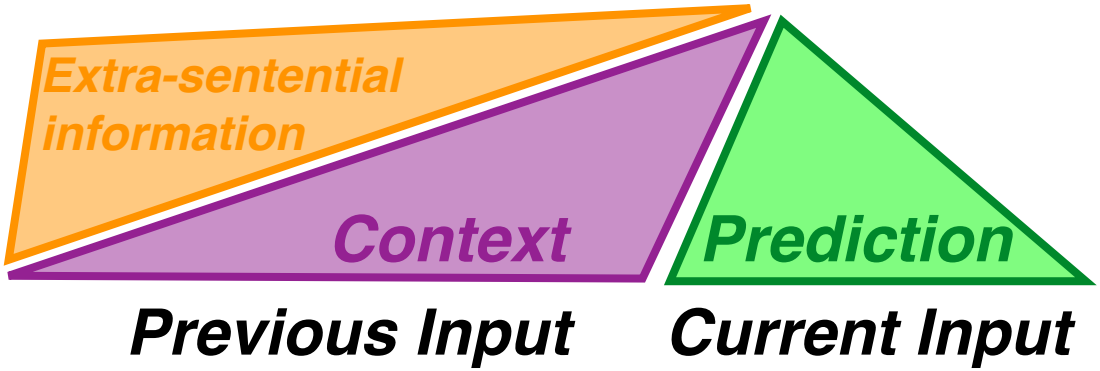
  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

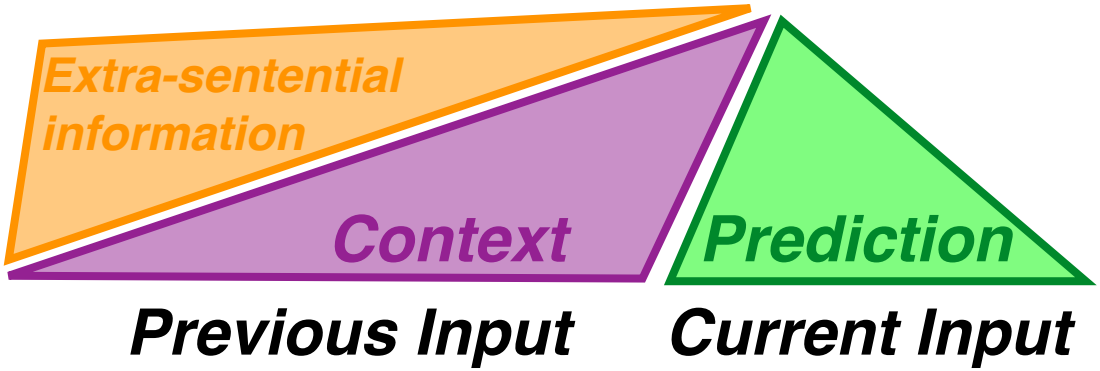
  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…

Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

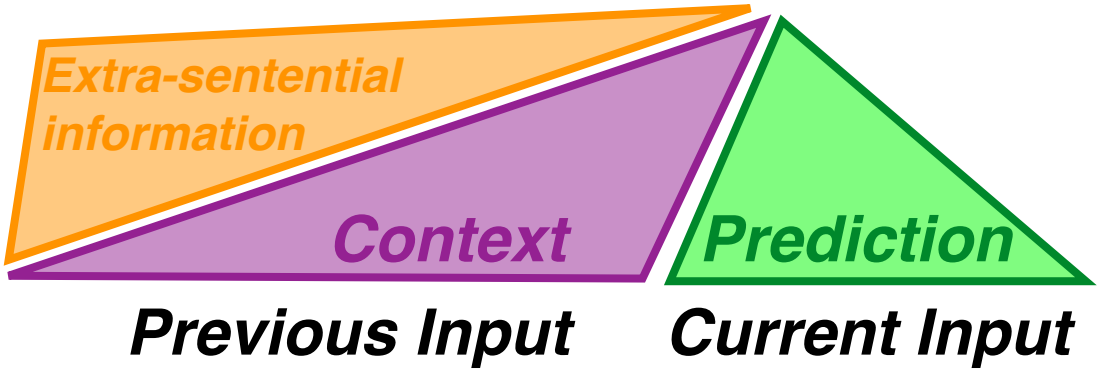
  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…play

Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

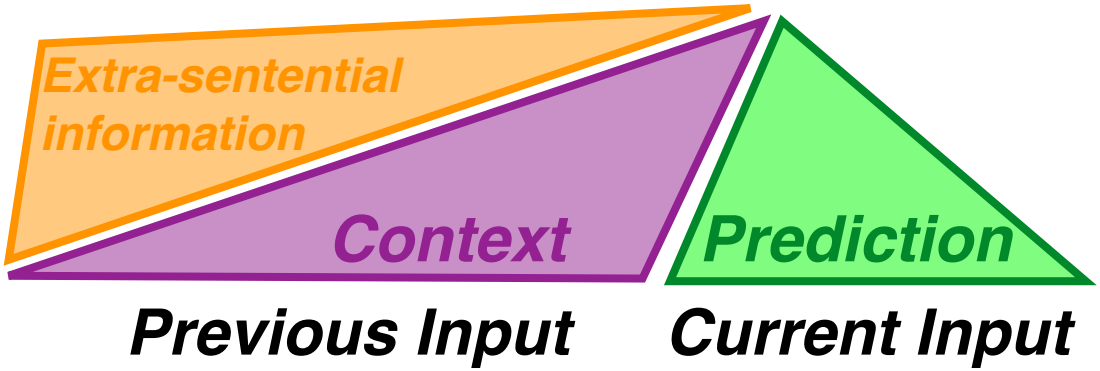
  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…play

  The squirrel stored some nuts in the…

Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

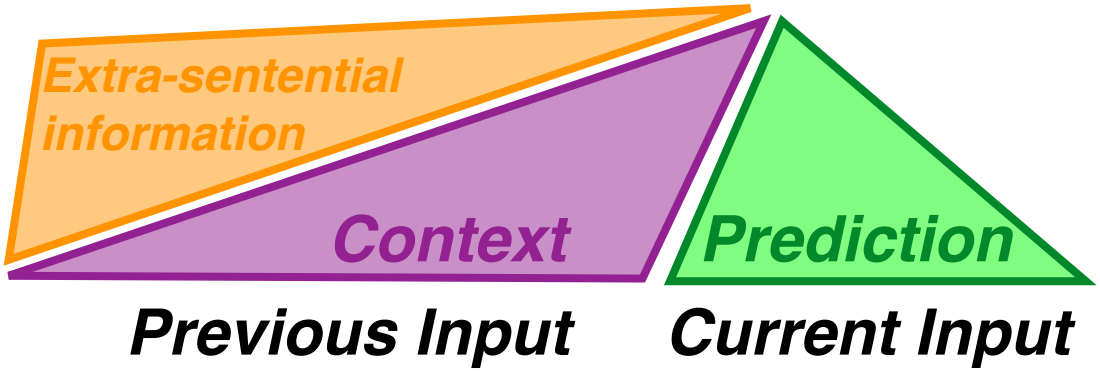
  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…play

  The squirrel stored some nuts in the…statue

Expectations in incremental comprehension



- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

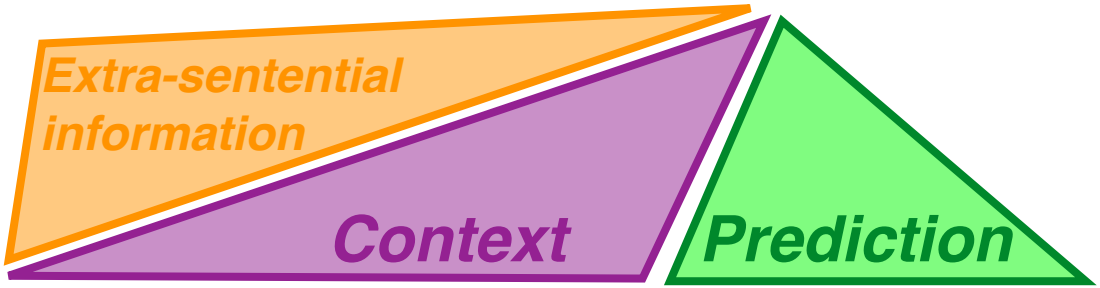
- Semantic & situational knowledge:

  The children went outside to…play

  The squirrel stored some nuts in the…~~statue~~

  tree

Expectations in incremental comprehension



*These expectations from **diverse contextual cues** affect human language processing **extremely quickly***

- Syntactic:

  Jamie was clearly intimidated… by [source]

- Phonological knowledge:

  Terry ate an… apple/orange/ice cream cone

  Terry ate a… nectarine/banana/sandwich

- Semantic & situational knowledge:

  The children went outside to…play

  The squirrel stored some nuts in the…~~statue~~

  tree

Surprisal as an index of real-time processing load

Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$

$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

(Hale, 2001, NAACL; Levy, 2008, Cognition)

Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$

$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process

(Hale, 2001, NAACL; Levy, 2008, Cognition)

Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$

$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process
  - Brains are prediction engines!

(Hale, 2001, NAACL; Levy, 2008, Cognition)

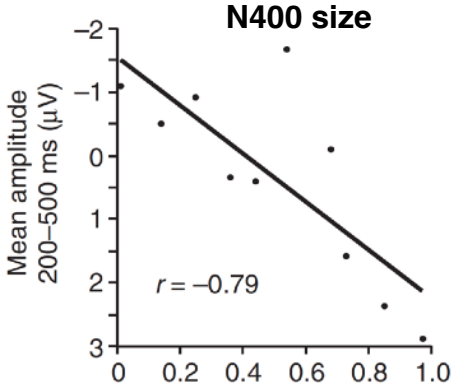
Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$

$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process
  - Brains are prediction engines!
- Predictable words are:
  - read faster (Ehrlich & Rayner, 1981)
  - have distinctive EEG responses (Kutas & Hillyard 1980)



(Federmeier & Kutas, 1999)

(Hale, 2001, NAACL; Levy, 2008, Cognition)

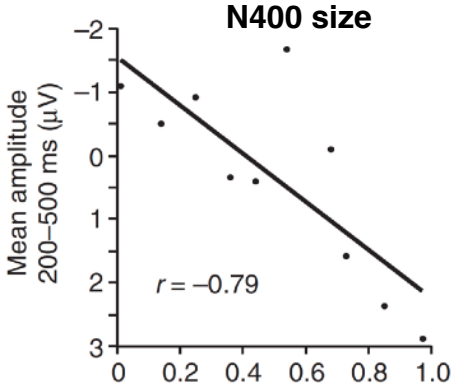
Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$

$$\left[\approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process
  - Brains are prediction engines!



(Federmeier & Kutas, 1999)

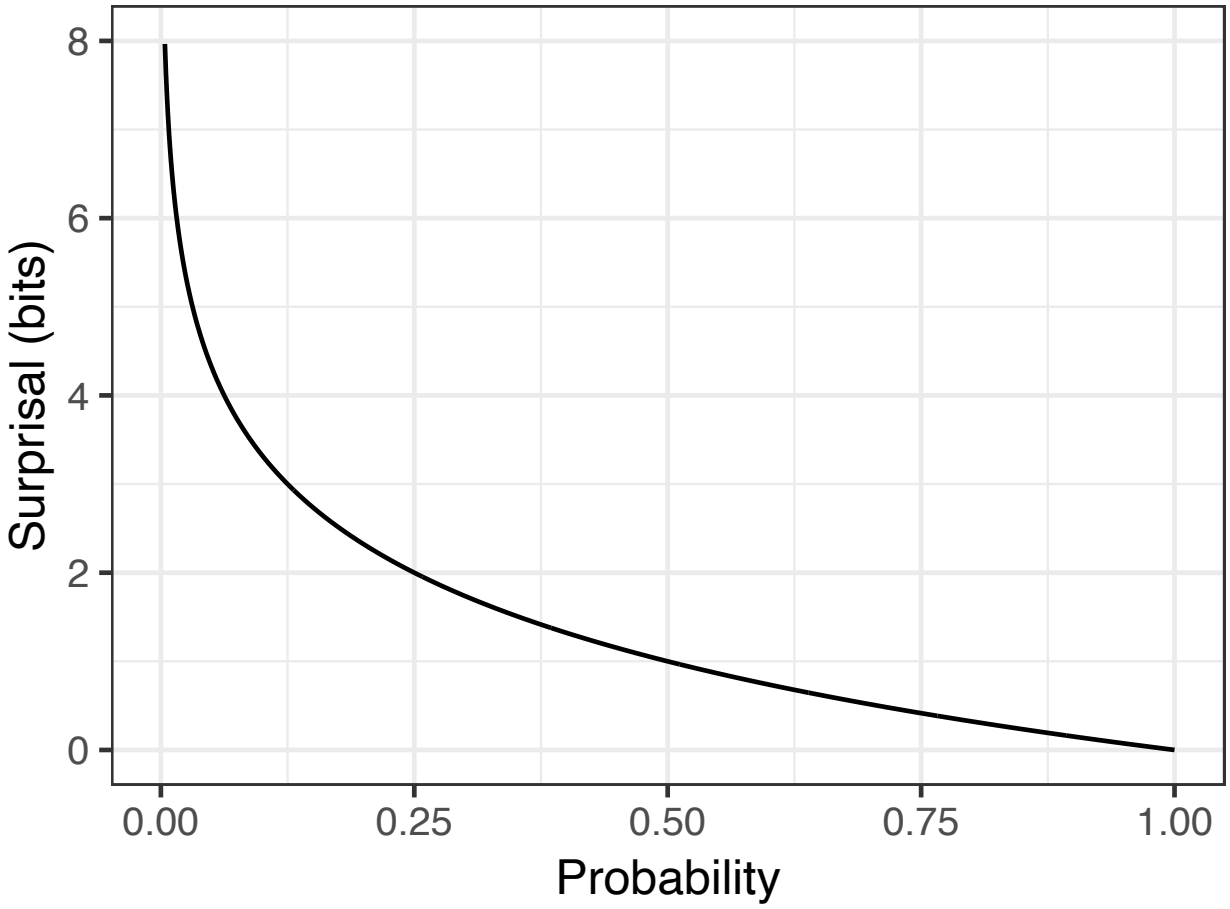
- Predictable words are:
  - read faster (Ehrlich & Rayner, 1981)
  - have distinctive EEG responses (Kutas & Hillyard 1980)
- with a language model that captures syntactic structure, we can get GRAMMATICAL EXPECTATIONS

(Hale, 2001, NAACL; Levy, 2008, Cognition)

Quantifying structure and surprise

- Hypothesis: a word's difficulty is its *surprisal* in context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$



(Shannon, 1948: a basic quantity from information theory!)

Estimating probability/time curve shape

Estimating probability/time curve shape

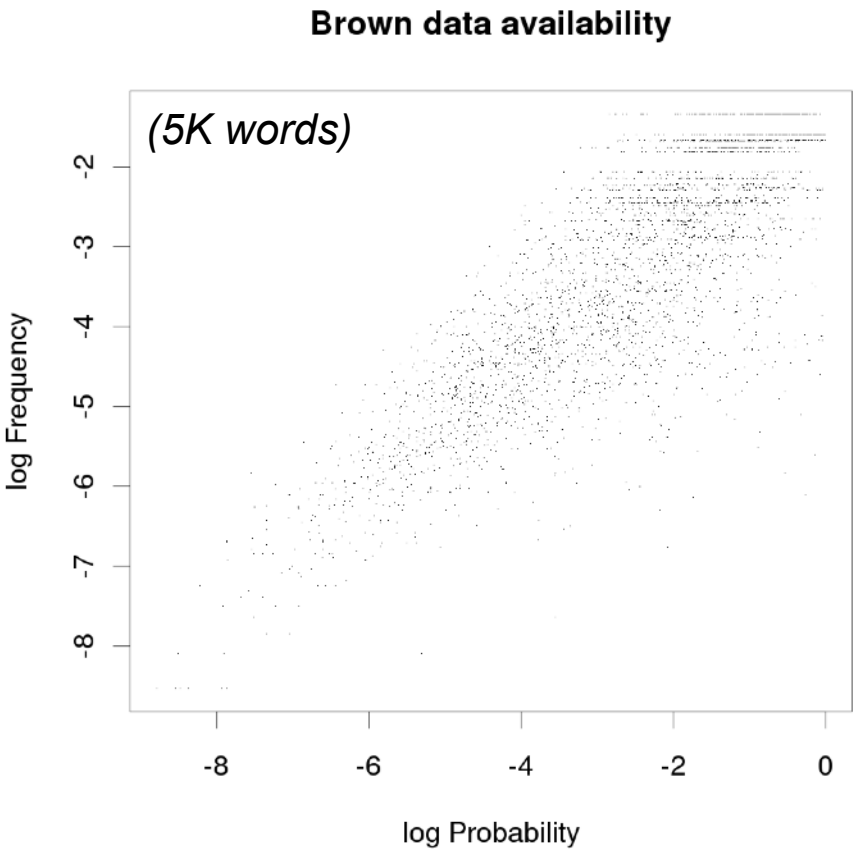
- As a proxy for “processing difficulty,” reading time in two different methods: self-paced reading & eye-tracking

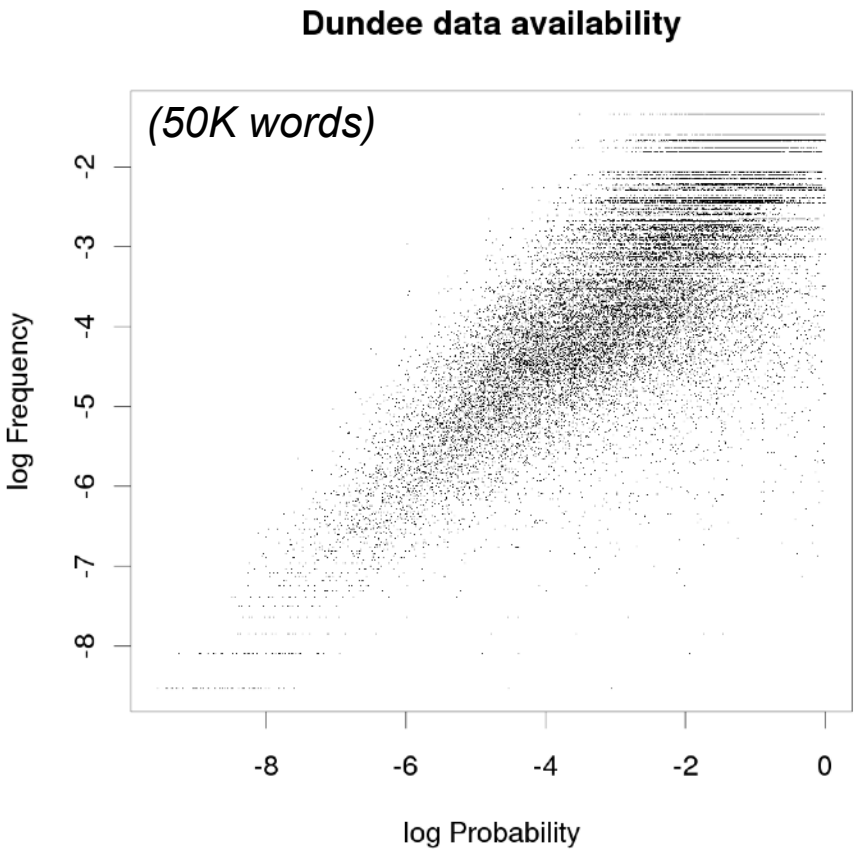
Estimating probability/time curve shape

- As a proxy for “processing difficulty,” reading time in two different methods: self-paced reading & eye-tracking
- Challenge: we need big data to estimate curve shape, but probability correlated with confounding variables

Estimating probability/time curve shape

- As a proxy for “processing difficulty,” reading time in two different methods: self-paced reading & eye-tracking
- Challenge: we need big data to estimate curve shape, but probability correlated with confounding variables

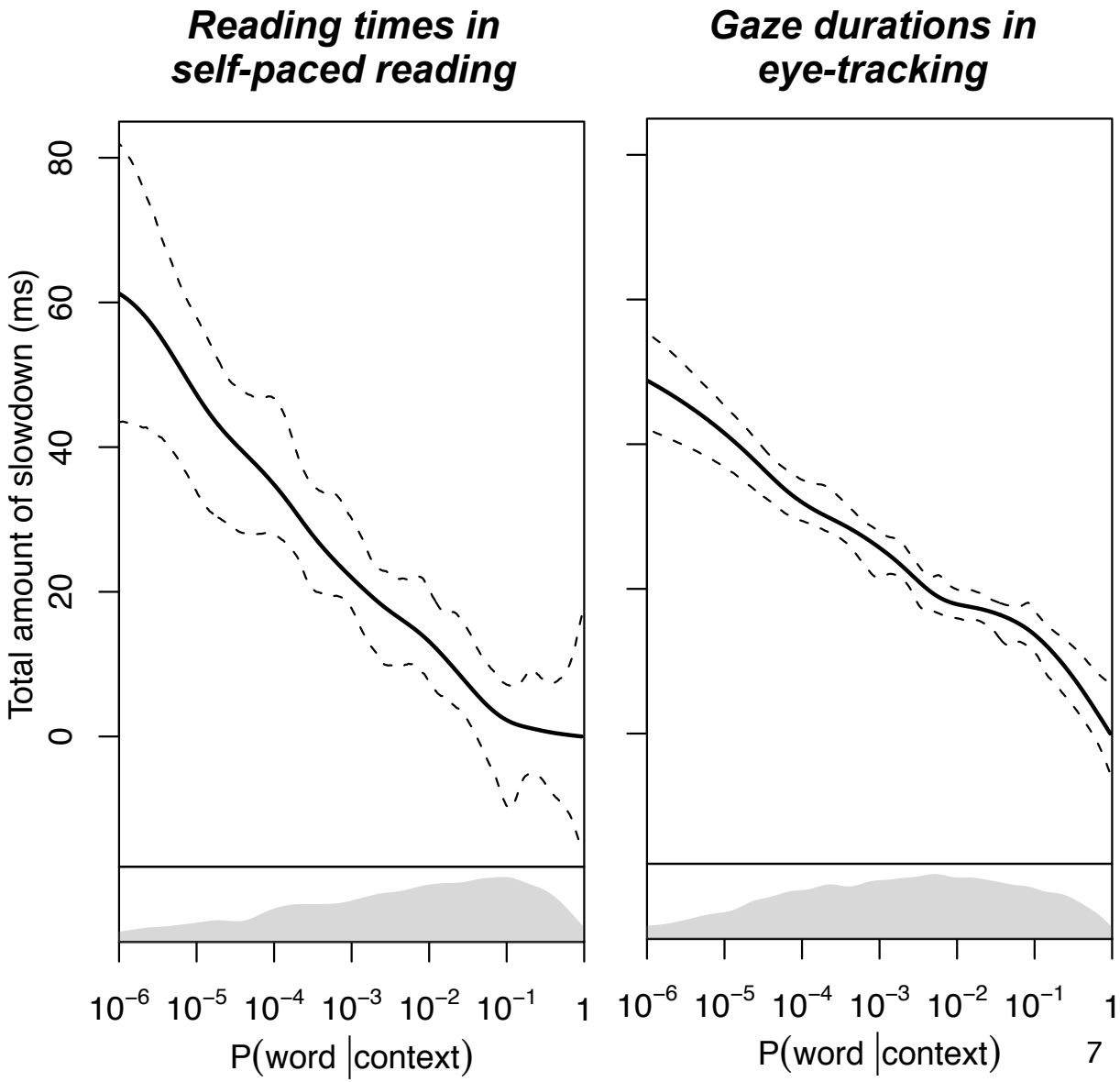




Estimating probability/time curve shape

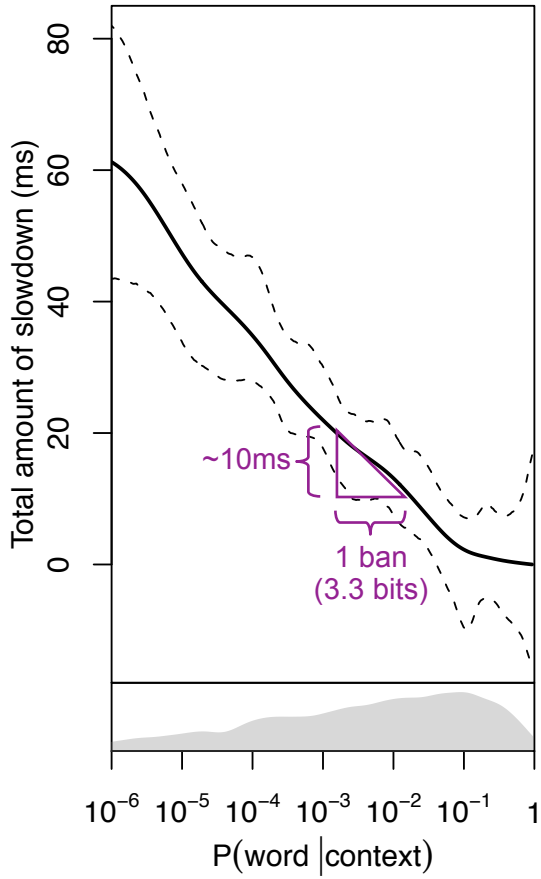
- Generalized additive model regression: total contribution of word (trigram) probability to RT near-linear over 6 orders of magnitude!

(Smith & Levy, 2013)



Take-away: how long to process a word in context?

- On average, time ***linear in the word's log-probability***
- Methodologically: reading puts control in the comprehender's hands (and eyes!), allowing us to study processing difficulty through reading time



A model system with incrementality, structure, and surprise

A model system with incrementality, structure, and surprise

The

A model system with incrementality, structure, and surprise

The woman

A model system with incrementality, structure, and surprise

The woman brought

A model system with incrementality, structure, and surprise

The woman brought the

A model system with incrementality, structure, and surprise

The woman brought the sandwich

A model system with incrementality, structure, and surprise

The woman brought the sandwich from

A model system with incrementality, structure, and surprise

The woman brought the sandwich from the

A model system with incrementality, structure, and surprise

The woman brought the sandwich from the kitchen

A model system with incrementality, structure, and surprise

The woman brought the sandwich from the kitchen tripped.

A model system with incrementality, structure, and surprise

A model system with incrementality, structure, and surprise

The woman who was given the sandwich from the kitchen tripped.

A model system with incrementality, structure, and surprise

The woman(who was given the sandwich from the kitchen)tripped.

A model system with incrementality, structure, and surprise

The woman((who was)given the sandwich from the kitchen)tripped.

A model system with incrementality, structure, and surprise

The woman(given the sandwich from the kitchen)tripped.

The woman((who was)given the sandwich from the kitchen)tripped.

A model system with incrementality, structure, and surprise

The woman((who was)brought the sandwich from the kitchen)tripped.

The woman(given the sandwich from the kitchen)tripped.

The woman((who was)given the sandwich from the kitchen)tripped.

A model system with incrementality, structure, and surprise

The woman(brought the sandwich from the kitchen)tripped.

The woman((who was)brought the sandwich from the kitchen)tripped.

The woman(given the sandwich from the kitchen)tripped.

The woman((who was)given the sandwich from the kitchen)tripped.

A model system with incrementality, structure, and surprise

The woman(brought the sandwich from the kitchen)tripped.

The woman((who was)brought the sandwich from the kitchen)tripped.

The woman(given the sandwich from the kitchen)tripped.

The woman((who was)given the sandwich from the kitchen)tripped.

	Simple past	Past participle
bring	***brought***	***brought***
give	***gave***	***given***

A model system with incrementality, structure, and surprise

The woman(brought the sandwich from the kitchen)tripped.

The woman((who was)brought the sandwich from the kitchen)tripped.

The woman(given the sandwich from the kitchen)tripped.

The woman((who was)given the sandwich from the kitchen)tripped.

	Simple past	Past participle
bring	***brought***	***brought***
give	***gave***	***given***

Low-tech, crowd-sourceable reading

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

F J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

of

F

dog

J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

pretty　　chased

F　　J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

the	eat
F	J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

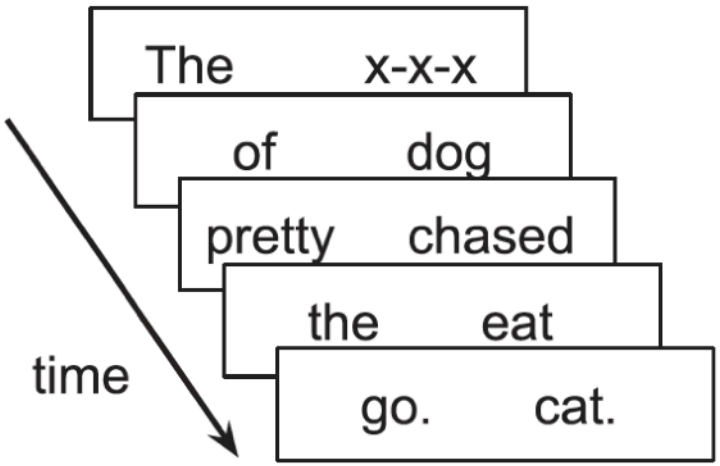
(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the word that fits given the preceding context

the eat

F J



(Forster et al., 2009; Boyce et al., 2020)

Incrementality, structure, and surprise

The woman brought the sandwich from the kitchen tripped.

The woman given the sandwich from the kitchen tripped.

The woman who was brought the sandwich from the kitchen tripped.

The woman who was given the sandwich from the kitchen tripped.

(Vani et al., in progress)

Incrementality, structure, and surprise

	Is the relative clause reduced?
The woman brought the sandwich from the kitchen tripped.	+
The woman given the sandwich from the kitchen tripped.	+
The woman who was brought the sandwich from the kitchen tripped.	–
The woman who was given the sandwich from the kitchen tripped.	–

(Vani et al., in progress)

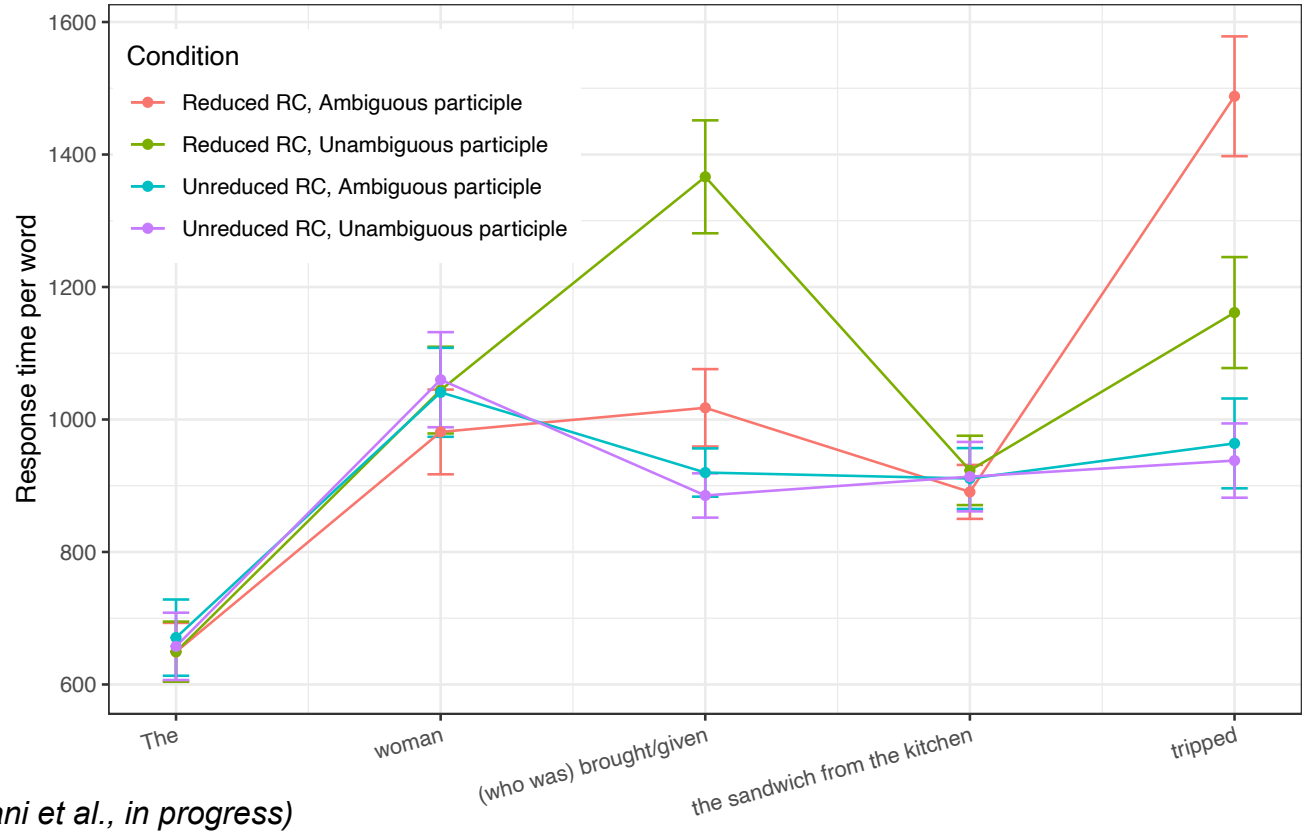
Incrementality, structure, and surprise

	Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	+	+
The woman given the sandwich from the kitchen tripped.	+	–
The woman who was brought the sandwich from the kitchen tripped.	–	+
The woman who was given the sandwich from the kitchen tripped.	–	–

(Vani et al., in progress)

Incrementality, structure, and surprise

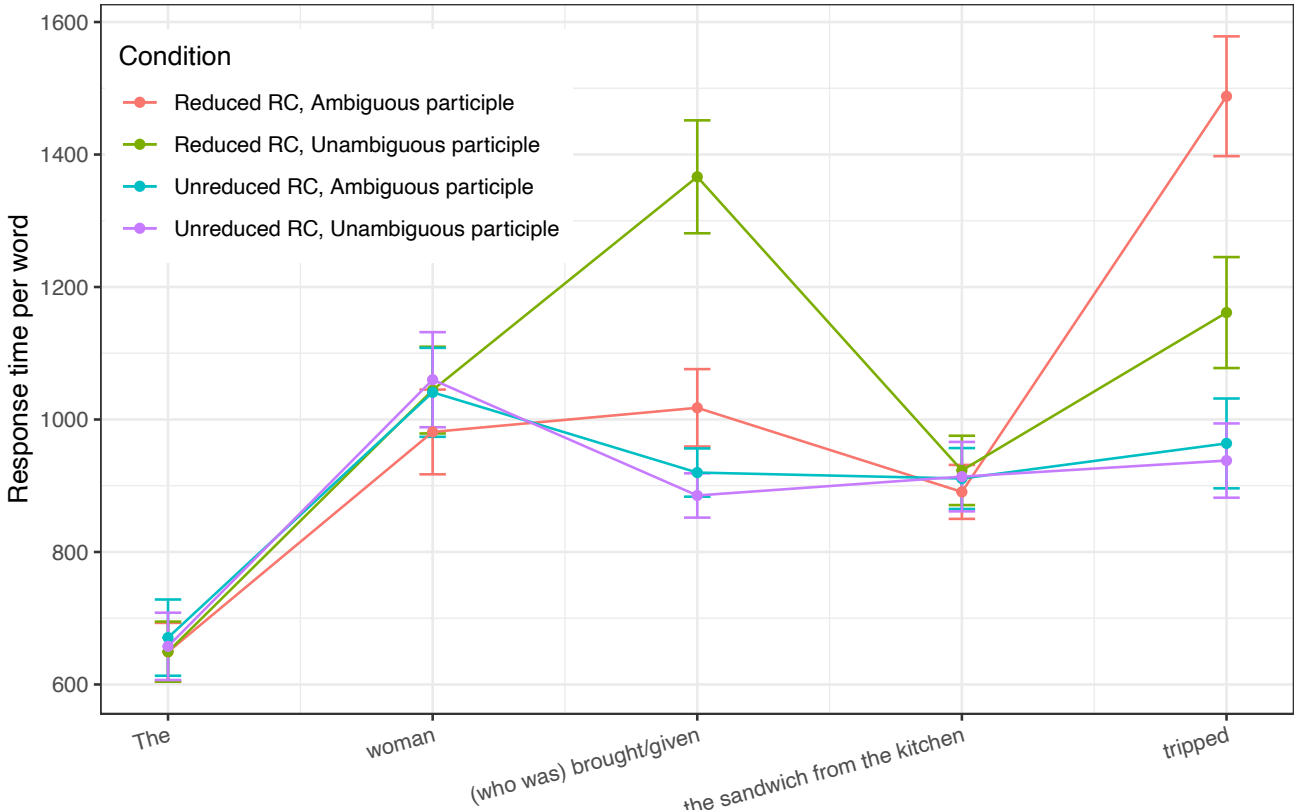
	Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	+	+
The woman given the sandwich from the kitchen tripped.	+	–
The woman who was brought the sandwich from the kitchen tripped.	–	+
The woman who was given the sandwich from the kitchen tripped.	–	–



(Vani et al., in progress)

Desiderata for human-like processing

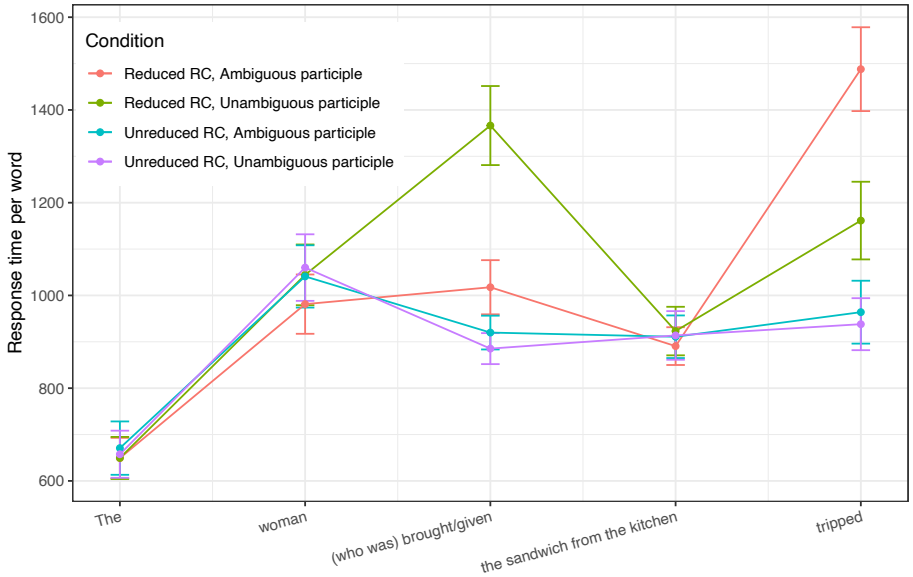
	Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	+	+
The woman given the sandwich from the kitchen tripped.	+	–
The woman who was brought the sandwich from the kitchen tripped.	–	+
The woman who was given the sandwich from the kitchen tripped.	–	–



(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

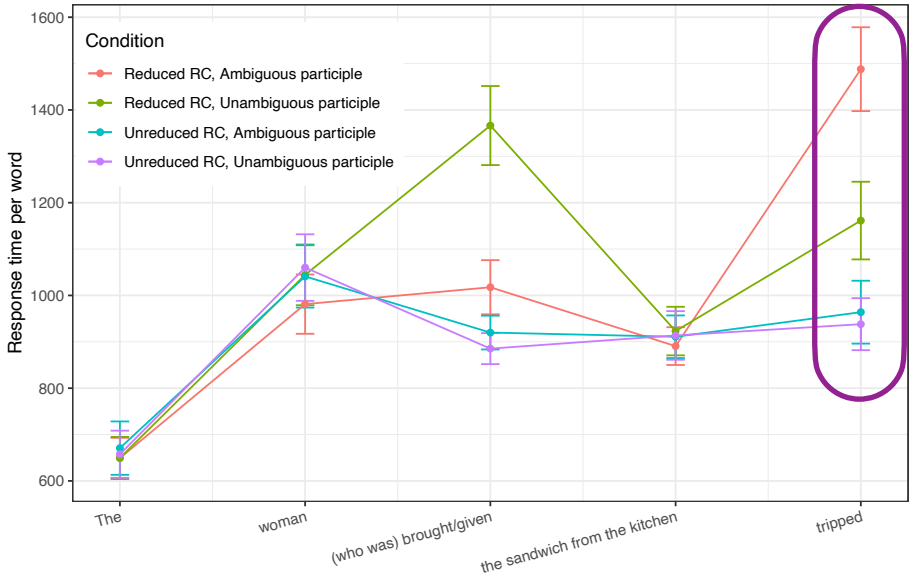
	Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	+	+
The woman given the sandwich from the kitchen tripped.	+	–
The woman who was brought the sandwich from the kitchen tripped.	–	+
The woman who was given the sandwich from the kitchen tripped.	–	–



(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

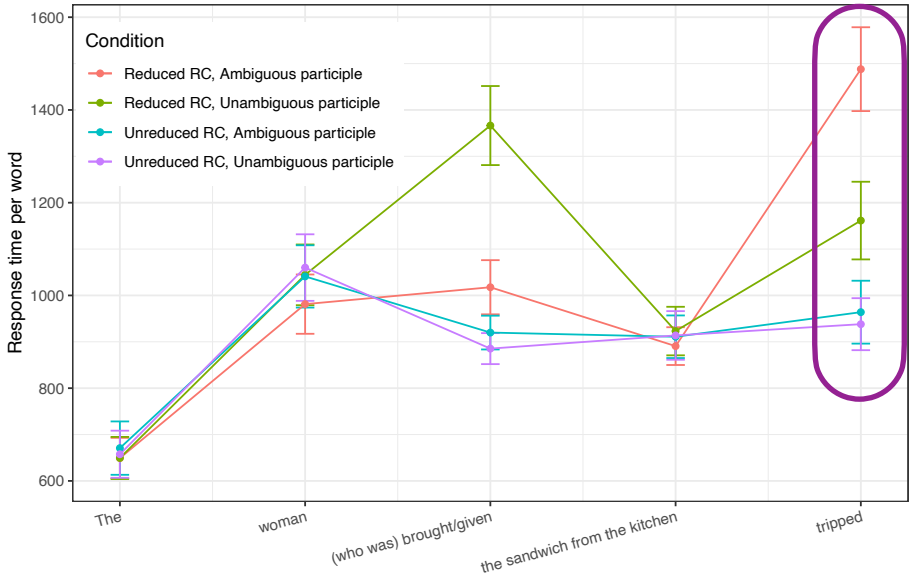
	Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	+	+
The woman given the sandwich from the kitchen tripped.	+	–
The woman who was brought the sandwich from the kitchen tripped.	–	+
The woman who was given the sandwich from the kitchen tripped.	–	–



(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

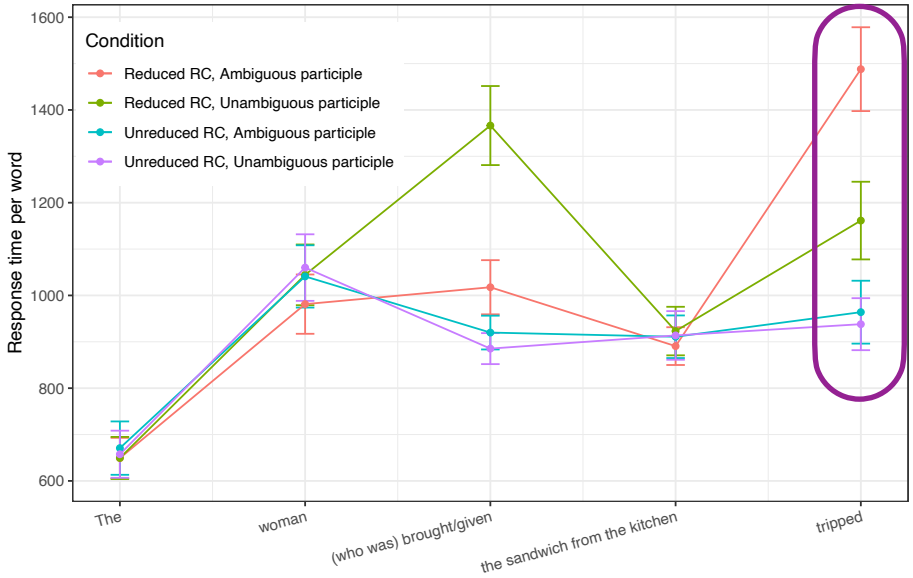
		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



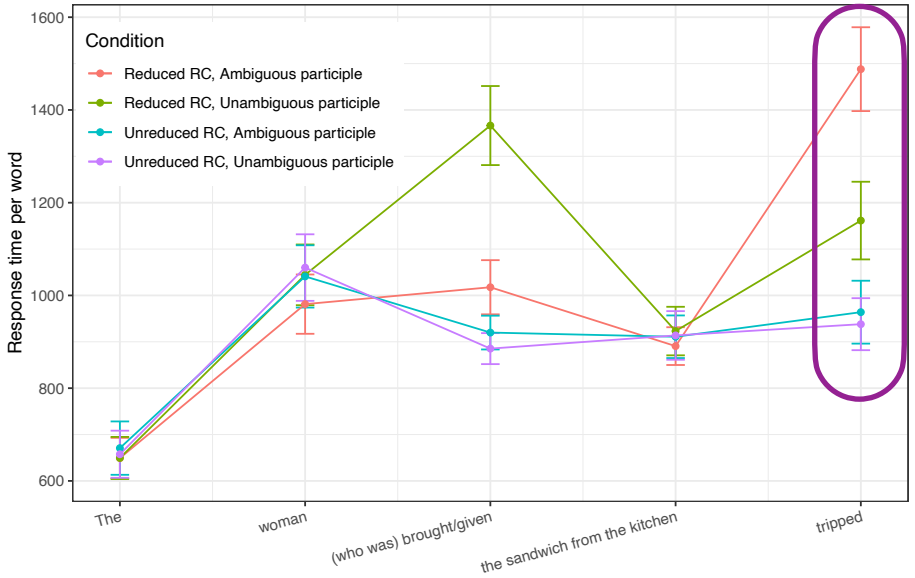
Define:

$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



Define:

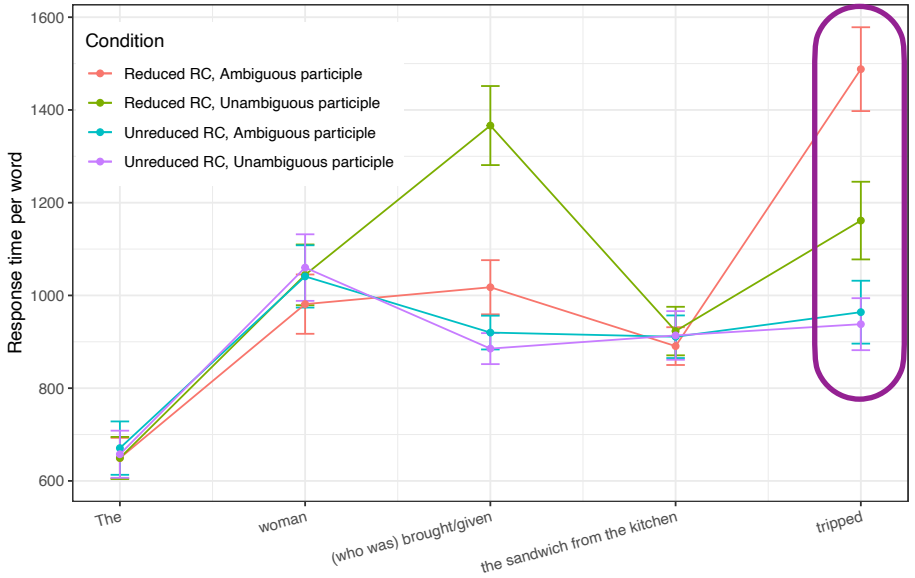
$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

Then we can define three criteria for "human-like" processing:

(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



Define:

$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

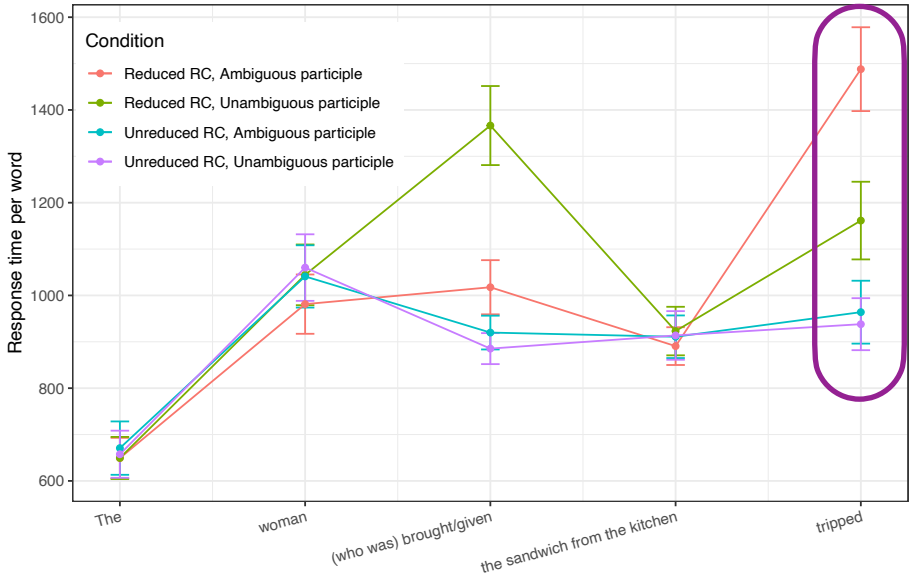
Then we can define three criteria for "human-like" processing:

(i) $S(a) > S(b)$

(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



Define:

$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

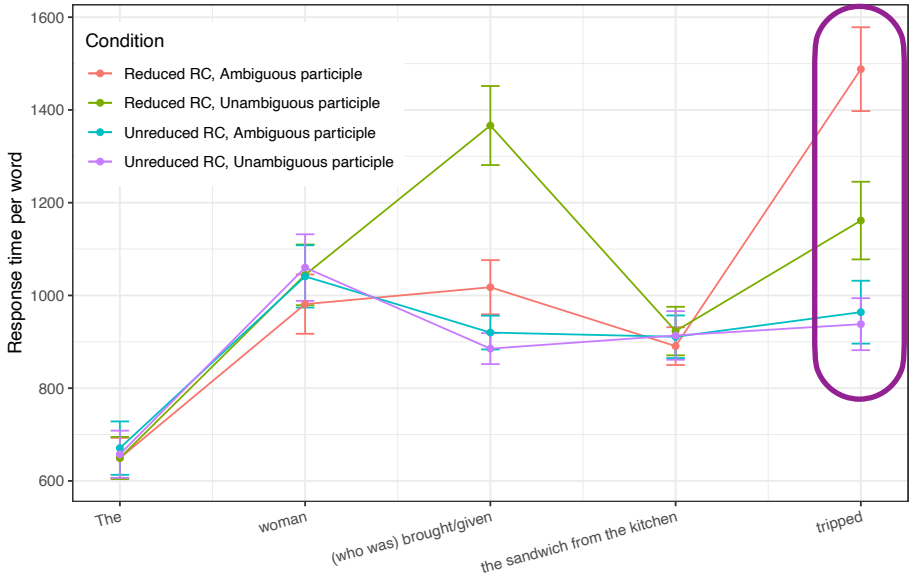
Then we can define three criteria for "human-like" processing:

(i) $S(a) > S(b)$
(ii) $S(a) > S(c)$

(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

Desiderata for human-like processing

		Is the relative clause reduced?	Is the participle part-of-speech ambiguous?
The woman brought the sandwich from the kitchen tripped.	a	+	+
The woman given the sandwich from the kitchen tripped.	b	+	–
The woman who was brought the sandwich from the kitchen tripped.	c	–	+
The woman who was given the sandwich from the kitchen tripped.	d	–	–



Define:

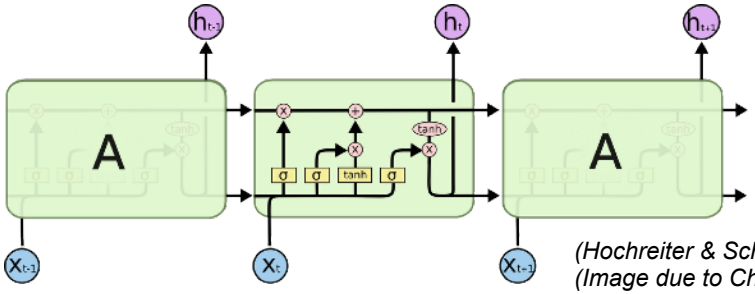
$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

Then we can define three criteria for "human-like" processing:

(i) $S(a) > S(b)$
(ii) $S(a) > S(c)$
(iii) $S(a) - S(b) > S(c) - S(d)$

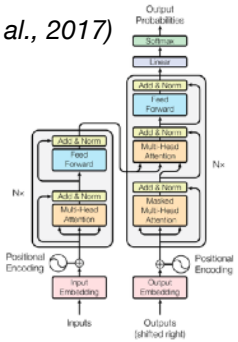
(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., in progress)

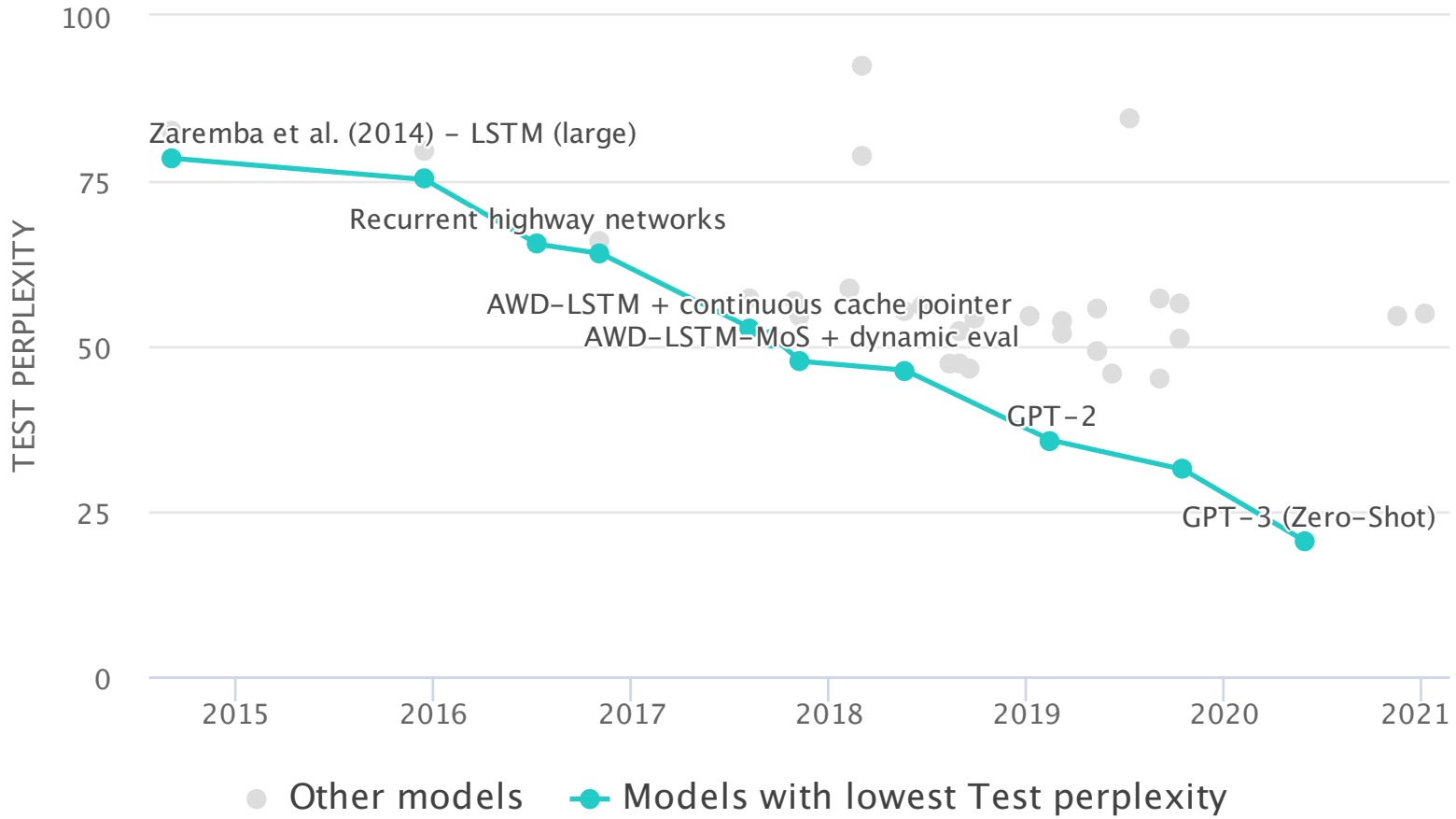
Deep learning has revolutionized language modeling



(Hochreiter & Schmidhuber, 1997)
(Image due to Christopher Olah)

(Vaswani et al., 2017)





https://paperswithcode.com/sota/language-modelling-on-penn-treebank-word

What have these models learned as “English”?

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper…

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful .

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry…

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded .

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

The man who the car…

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

The man who the car has gazed longingly at for years .

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

The man who the car has gazed longingly at for years . ✗

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

The man who the car has gazed longingly at for years . ✗

The athlete who the restaurant…

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

The man who the car has gazed longingly at for years . ✗

The athlete who the restaurant would justify decided to add the main West Coast restaurants to his menu and who hadn 't upgraded from his previous suite , into a more <UNK> steakhouse in New York .

(continuations generated by model of Jozefowicz et al., 2016)

What have these models learned as “English”?

The girl who the newspaper now calls his girlfriend has really been hateful . ✓

The monologue that the actor who the movie industry likes made silent was being uploaded . ✓ (!)

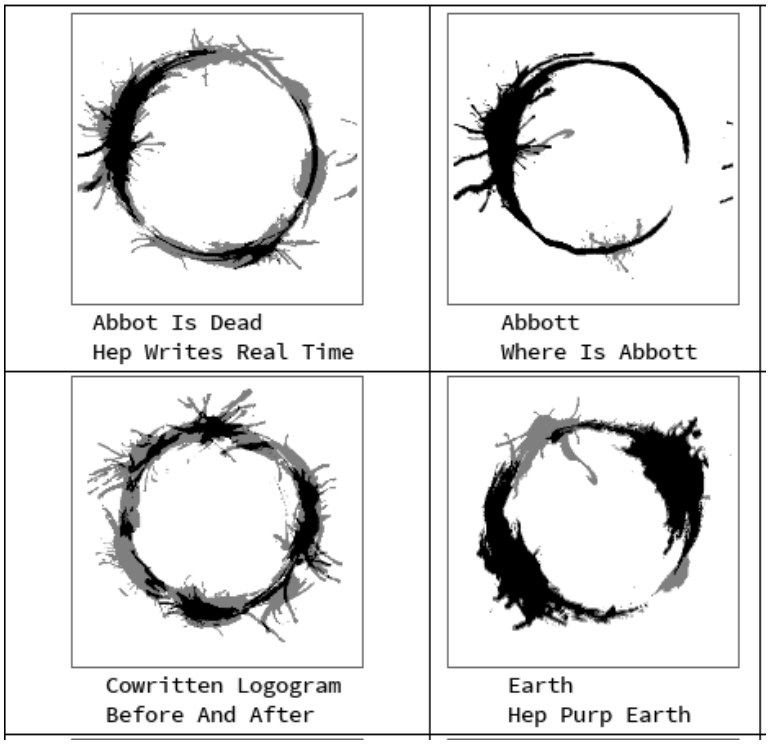
The man who the car has gazed longingly at for years . ✗

The athlete who the restaurant would justify decided to add the main West Coast restaurants to his menu and who hadn 't upgraded from his previous suite , into a more <UNK> steakhouse in New York . ???

(continuations generated by model of Jozefowicz et al., 2016)

Technical question:

What generalizations are these models learning?



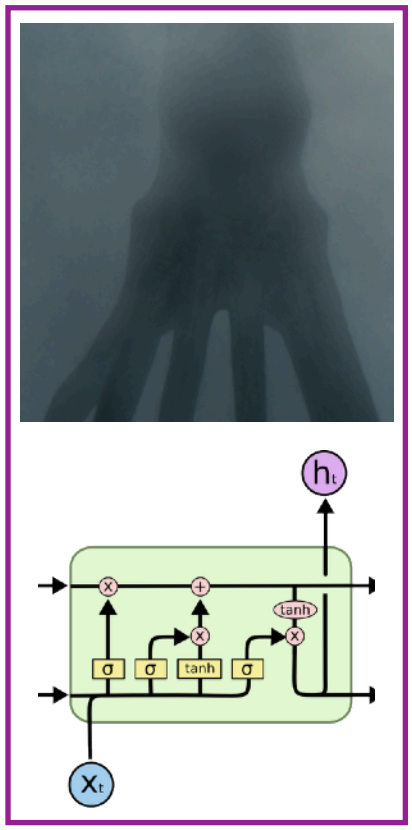
Theoretical question:

How well would positive* input data alone deliver the right linguistic generalizations to a generic flexible learner without strong hierarchical bias?

******No negative evidence!***

Theoretical question:

How well would positive* input data alone deliver the right linguistic generalizations to a generic flexible learner without strong hierarchical bias?



***No negative evidence!**

Focus: *grammatical generalizations*

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:
I bought the bed

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:	Japanese:		
I bought the bed	beddo	-o	ka-tta
	(pro) bed	-ACC	buy-PAST

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:

I bought the bed

Japanese:

beddo -o ka-tta
(pro) bed -ACC buy-PAST

Oneida (Baker, 1996):

Wa' -ke -nakt -a -hnínu -'
FACT -1sS -bed -∅ -buy -PUNC

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:
I bought the bed

Japanese:
beddo -o ka-tta
(pro) bed -ACC buy-PAST

Oneida (Baker, 1996):
Wa' -ke -nakt -a -hnínu -'
FACT -1sS -bed -∅ -buy -PUNC

- Yet there are strong (universal?) generalizations

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:
I bought the bed

Japanese:
beddo -o ka-tta
(pro) bed -ACC buy-PAST

Oneida (Baker, 1996):
Wa' -ke -nakt -a -hnínu -'
FACT -1sS -bed -∅ -buy -PUNC

- Yet there are strong (universal?) generalizations

Grammatical categories:

N V Adj Prep

Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:
I bought the bed

Japanese:
beddo -o ka-tta
(pro) bed -ACC buy-PAST

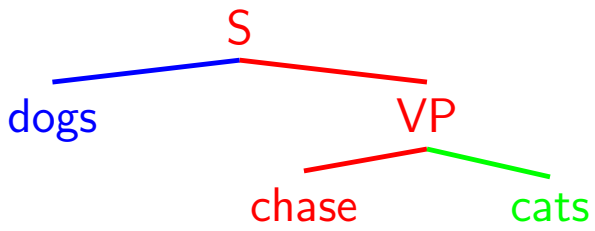
Oneida (Baker, 1996):
Wa' -ke -nakt -a -hnínu -'
FACT -1sS -bed -∅ -buy -PUNC

- Yet there are strong (universal?) generalizations

Grammatical categories:

N V Adj Prep

Heads & hierarchy:



Focus: *grammatical generalizations*

- Languages vary dramatically across the world in structure

English:

I bought the bed

Japanese:

beddo -o ka-tta
(pro) bed -ACC buy-PAST

Oneida (Baker, 1996):

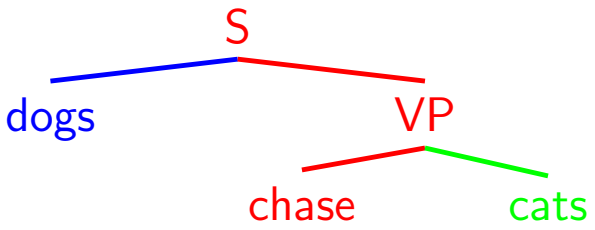
Wa' -ke -nakt -a -hnínu -'
FACT -1sS -bed -∅ -buy -PUNC

- Yet there are strong (universal?) generalizations

Grammatical categories:

N V Adj Prep

Heads & hierarchy:



Idiosyncrasy:

$[\![\text{kick the bucket}]\!]$

$\neq$

$[\![\text{kick}]\!](\iota(\lambda x.[\![bucket]\!](x)))$

Focus: *grammatical generalizations*

INFORMATION AND CONTROL 10, 447–474 (1967)

Language Identification in the Limit

E MARK GOLD*

The RAND Corporation

Language learnability has been investigated. This refers to the following situation: A class of possible languages is specified, together with a method of presenting information to the learner about an unknown language, which is to be chosen from the class. The question is now asked, "Is the information sufficient to determine which of the possible languages is the unknown language?" Many definitions of learnability are possible, but only the following is considered here: Time is quantized and has a finite starting time. At each time the learner receives a unit of information and is to make a guess as to the identity of the unknown language on the basis of the information received so far. This process continues forever. The class of languages will be considered *learnable* with respect to the specified method of information presentation if there is an algorithm that the learner can use to make his guesses, the algorithm having the following property: Given any language of the class, there is some finite time after which the guesses will all be the same and they will be correct.

Linguistic Inquiry Volume 10 Number 4 (Fall, 1979) 533–581.

C. L. Baker

Syntactic Theory and the Projection Problem*

0. Introduction

One of the most basic concerns in the writings of Noam Chomsky, beginning in *Syntactic Structures* and extending through his recent work, has been to draw attention to the profundity of one of the central psychological problems posed by the phenomenon of first language acquisition.[1] Following Peters (1972), I will refer to this problem as the "projection problem". Reduced to essentials, solving the projection problem involves finding a satisfactory answer to the following question: What is the functional relation that exists between an arbitrary human being's early linguistic experience (his "primary linguistic data") and his resulting adult intuitions?[2] A solution to this problem requires a body of hypotheses that would make it possible to deduce the full range of adult intuitions in advance, given only a suitable record of the early experience. Applied to

Linguistic Nativism and the Poverty of the Stimulus

Alexander Clark and Shalom Lappin

Focus on learning of *grammar*

Focus on learning of *grammar*

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***

Focus on learning of *grammar*

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)

Focus on learning of *grammar*

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)
  - However, these models' predictions are **not quantitatively aligned with human comprehension behavior when expectations about grammatical structure are violated**

Focus on learning of *grammar*

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)
  - However, these models' predictions are **not quantitatively aligned with human comprehension behavior when expectations about grammatical structure are violated**
  - Deep-learning models offer **insights into learnability** and **a powerful scientific tool for expectation estimation,** but **not a theoretical account** of human language representation and processing

My strategy and argument today

My strategy and argument today

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***

My strategy and argument today

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)

My strategy and argument today

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)
  - However, these models' predictions are **not quantitatively aligned with human comprehension behavior when expectations about grammatical structure are violated**

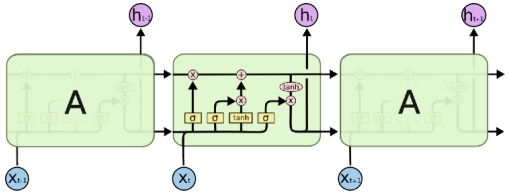
My strategy and argument today

- **Strategy:** run ***controlled experiments*** on models as if they were ***subjects in a psycholinguistics experiment***
- **Conclusions:**
  - Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)
  - However, these models' predictions are **not quantitatively aligned with human comprehension behavior when expectations about grammatical structure are violated**
  - Deep-learning models offer **insights into learnability** and **a powerful scientific tool for expectation estimation,** but **not a theoretical account** of human language representation and processing

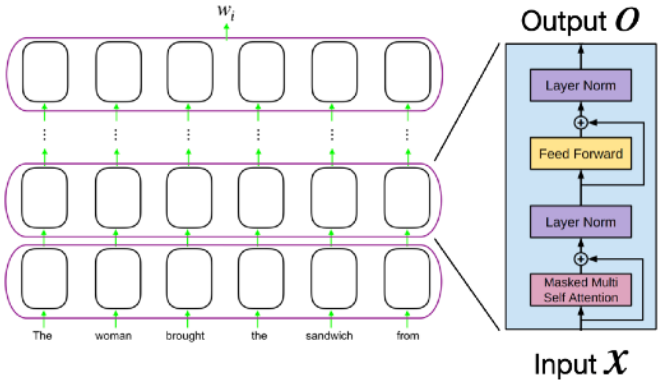
Language models tested

Model	Architecture	Training Data	Data size (tokens)	Reference
JRNN	LSTM	Web text	~800,000,000	Jozefowicz et al. (2016)
GRNN	LSTM	Wikipedia	~90,000,000	Gulordava et al. (2018)
GPT-2	Transformer	Web text	~8,000,000,000	Radford et al. (2019)
GPT-3	Transformer	Web text	~40,000,000,000	Brown et al. (2020)
RNNG	Syntax+LSTM	Penn Treebank	~1,000,000	Dyer et al. (2016)
tinylstm	LSTM	Penn Treebank	~1,000,000	
n-gram	5-gram model	British Nat'l Corpus	~100,000,000	

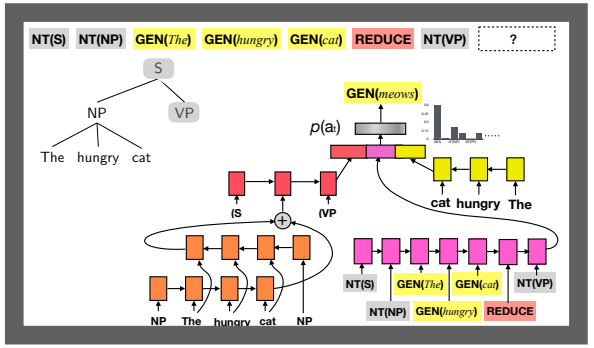
Long Short-Term Memory (LSTM) Network



Transformer decoder



Recurrent Neural Network Grammar

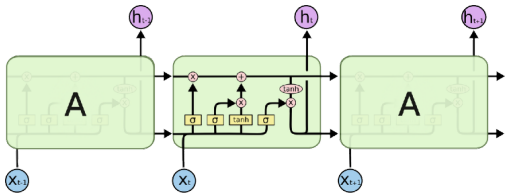


Language models tested

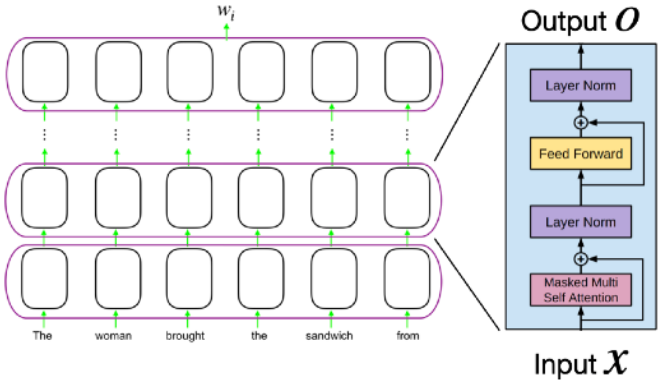
*~human **childhood** of linguistic experience*

Model	Architecture	Training Data	Data size (tokens)	Reference
JRNN	LSTM	Web text	~800,000,000	Jozefowicz et al. (2016)
GRNN	LSTM	Wikipedia	~90,000,000	Gulordava et al. (2018)
GPT-2	Transformer	Web text	~8,000,000,000	Radford et al. (2019)
GPT-3	Transformer	Web text	~40,000,000,000	Brown et al. (2020)
RNNG	Syntax+LSTM	Penn Treebank	~1,000,000	Dyer et al. (2016)
tinylstm	LSTM	Penn Treebank	~1,000,000	
n-gram	5-gram model	British Nat'l Corpus	~100,000,000	

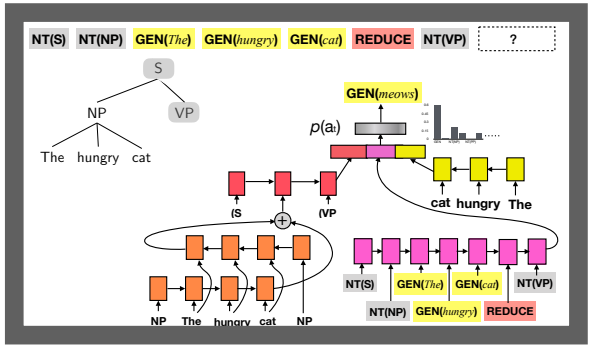
Long Short-Term Memory (LSTM) Network



Transformer decoder



Recurrent Neural Network Grammar



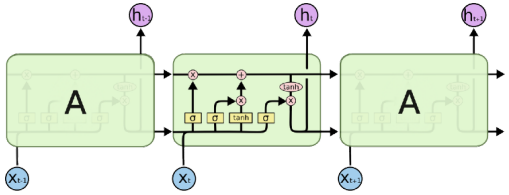
Language models tested

*~human **lifetime** of linguistic experience*

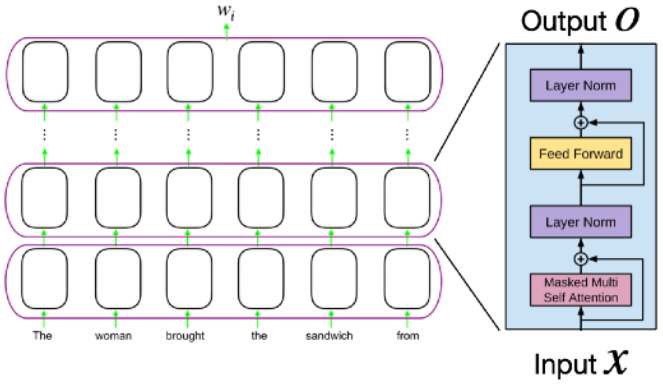
*~human **childhood** of linguistic experience*

Model	Architecture	Training Data	Data size (tokens)	Reference
JRNN	LSTM	Web text	~800,000,000	Jozefowicz et al. (2016)
GRNN	LSTM	Wikipedia	~90,000,000	Gulordava et al. (2018)
GPT-2	Transformer	Web text	~8,000,000,000	Radford et al. (2019)
GPT-3	Transformer	Web text	~40,000,000,000	Brown et al. (2020)
RNNG	Syntax+LSTM	Penn Treebank	~1,000,000	Dyer et al. (2016)
tinylstm	LSTM	Penn Treebank	~1,000,000	
n-gram	5-gram model	British Nat'l Corpus	~100,000,000	

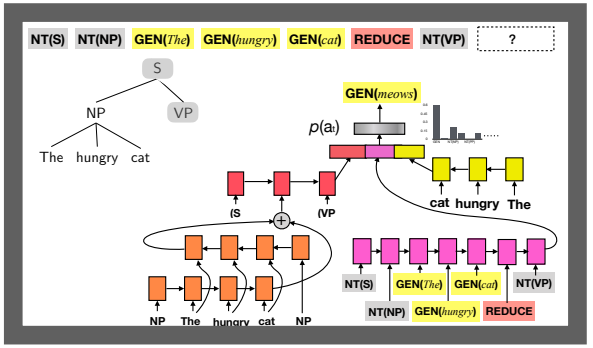
Long Short-Term Memory (LSTM) Network



Transformer decoder



Recurrent Neural Network Grammar



Simplest syntactic hierarchy: subordination

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

The doctor studied the textbook .

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

As the doctor studied the textbook .

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

The doctor studied the textbook
, the nurse walked into the office .

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

As the doctor studied the textbook , the nurse walked into the office .

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

Context

✓ *The doctor studied the textbook* | *.*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook*
, the nurse walked into the office .

✓ *As the doctor studied the textbook*
, the nurse walked into the office .

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

Context | Completion

✓ *The doctor studied the textbook* | *.*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$$-\log P(\text{Completion}|\text{Context})$$

Context | Completion

✓ *The doctor studied the textbook* | *.*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$-\log P(\text{Completion}|\text{Context})$

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$-\log P(\text{Completion}|\text{Context})$

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$-\log P(\text{Completion}|\text{Context})$

"No-matrix" variants
(No subsequent matrix clause)

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$-\log P(\text{Completion}|\text{Context})$

"No-matrix" variants
(No subsequent matrix clause)

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

"Matrix" variants
(There is a subsequent matrix clause)

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$$-\log P(\text{Completion}|\text{Context})$$

"No-matrix" variants
(No subsequent matrix clause)

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

Surprisal difference (should be *positive*)

"Matrix" variants
(There is a subsequent matrix clause)

? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Simplest syntactic hierarchy: subordination

$$-\log P(\text{Completion}|\text{Context})$$

"No-matrix" variants
(No subsequent matrix clause)

✓ *The doctor studied the textbook .*

✗ *As the doctor studied the textbook .*

Surprisal difference (should be *positive*)

"Matrix" variants
(There is a subsequent matrix clause)

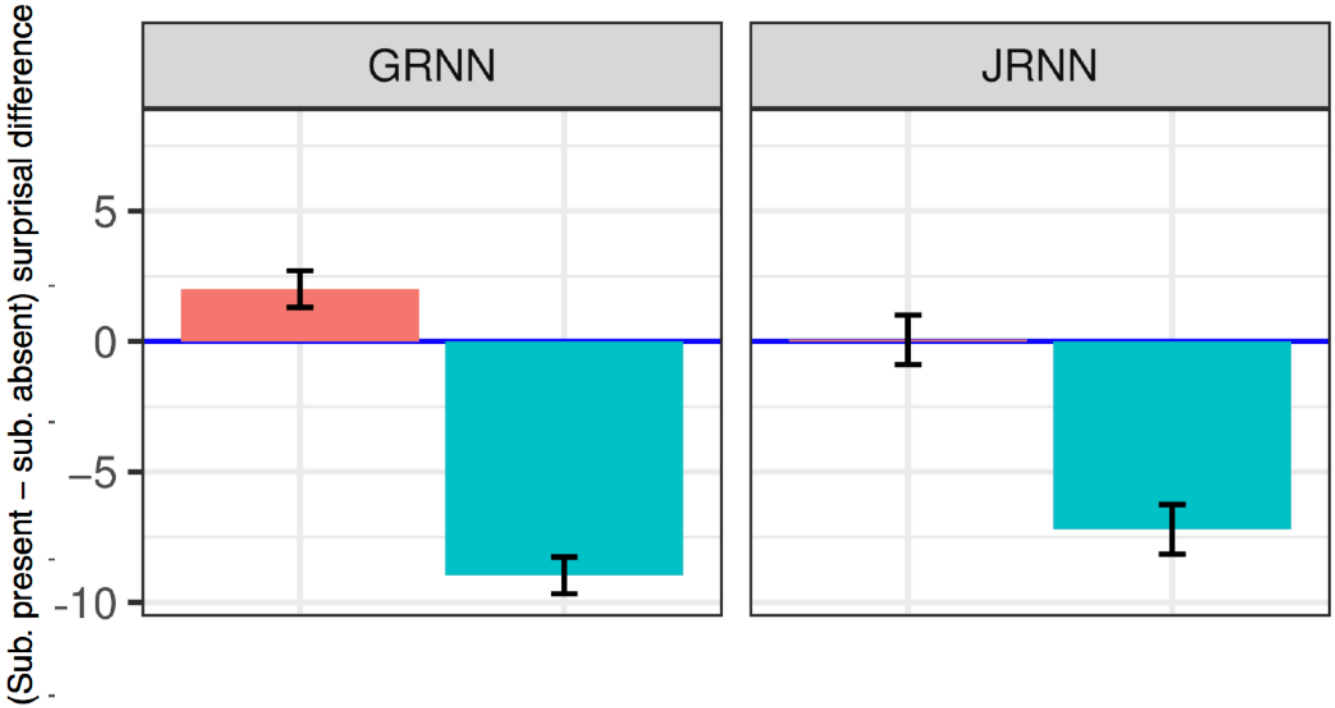
? *The doctor studied the textbook , the nurse walked into the office .*

✓ *As the doctor studied the textbook , the nurse walked into the office .*

Surprisal difference (should be *negative*)

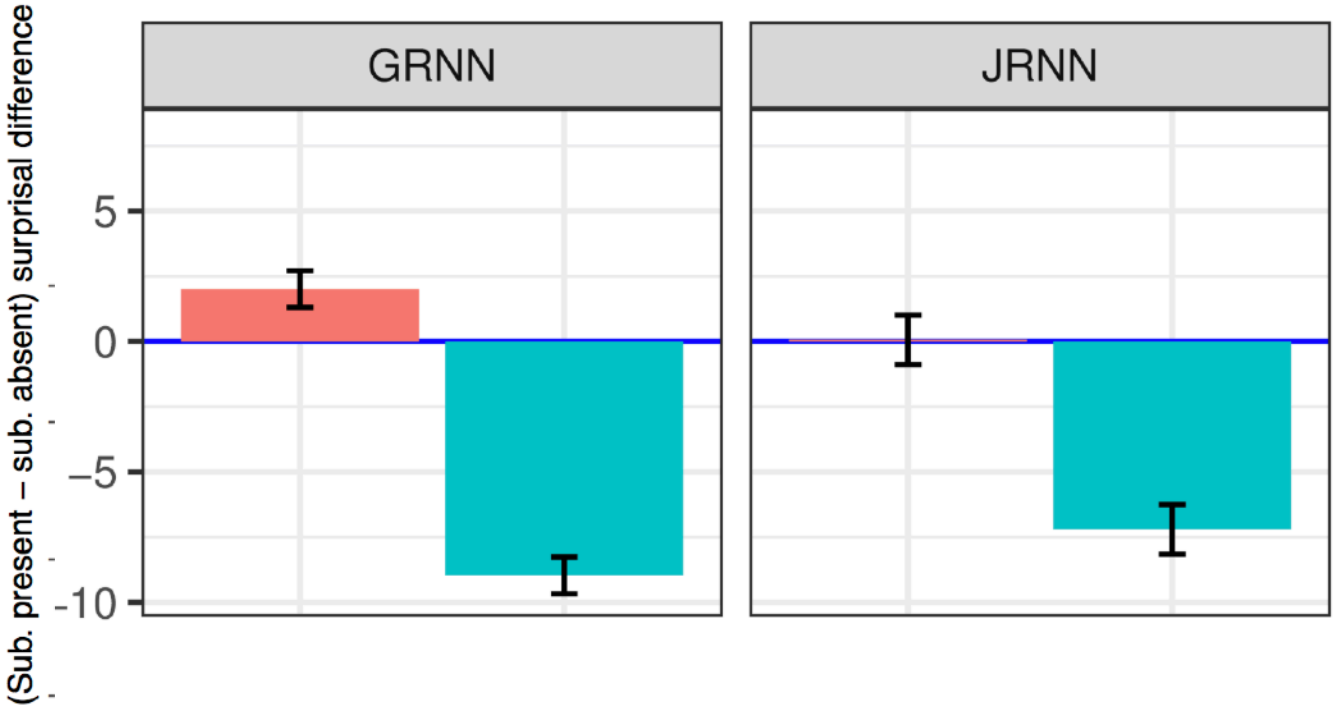
(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Effects of Subordinate Clauses



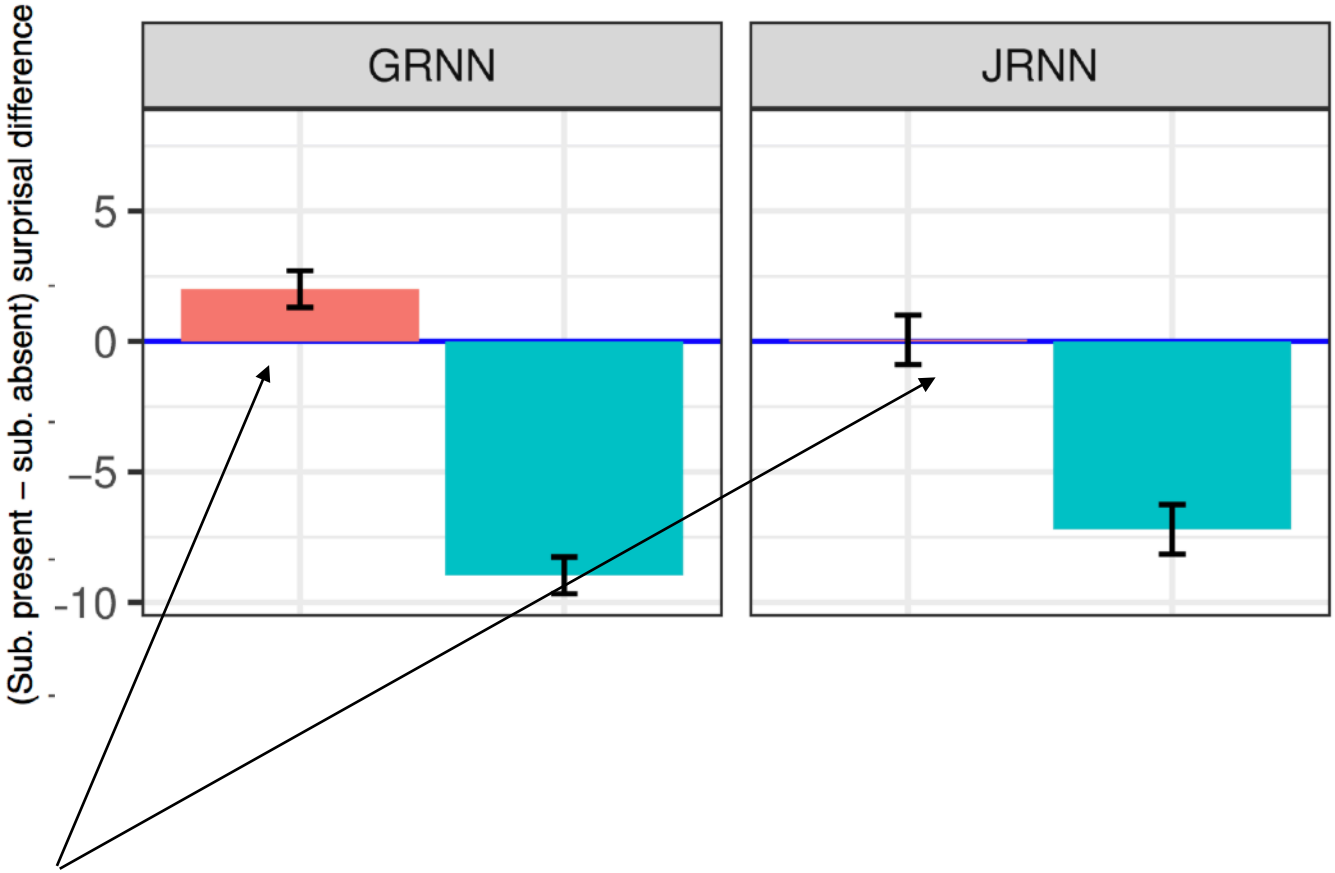
(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Effects of Subordinate Clauses



(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

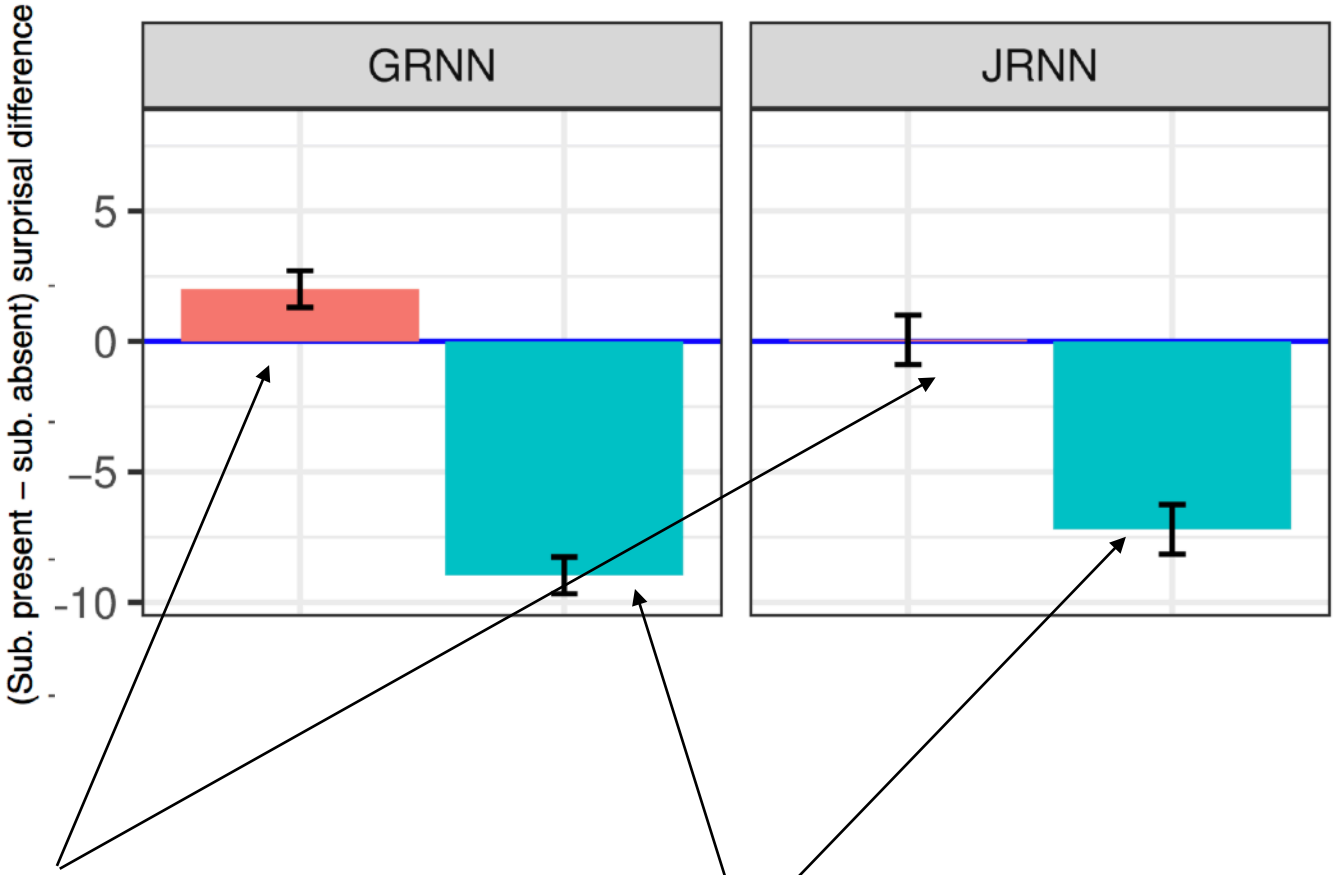
Effects of Subordinate Clauses



No-matrix penalty

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Effects of Subordinate Clauses

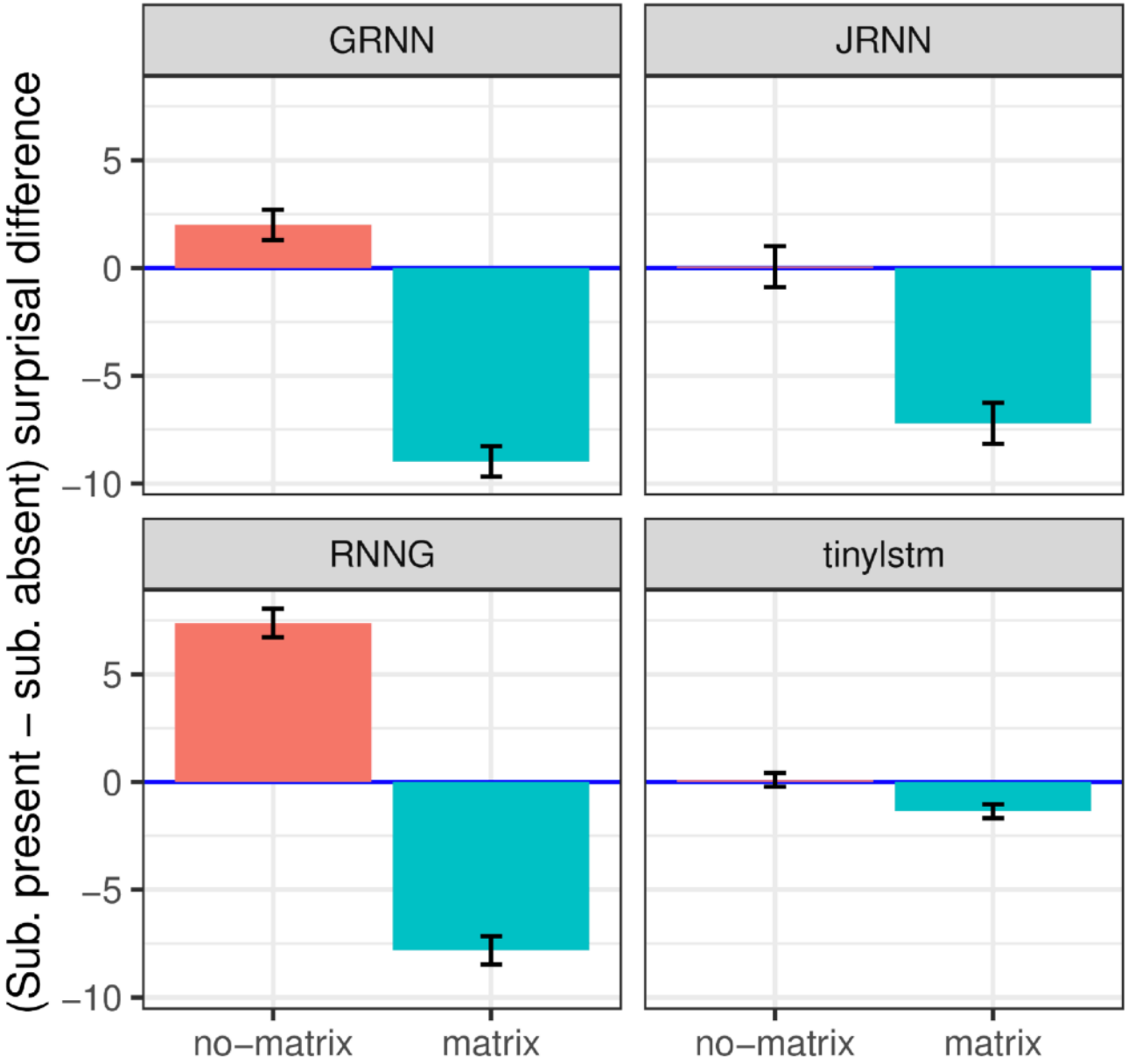


No-matrix penalty

Matrix licensing

(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

Subordination: results

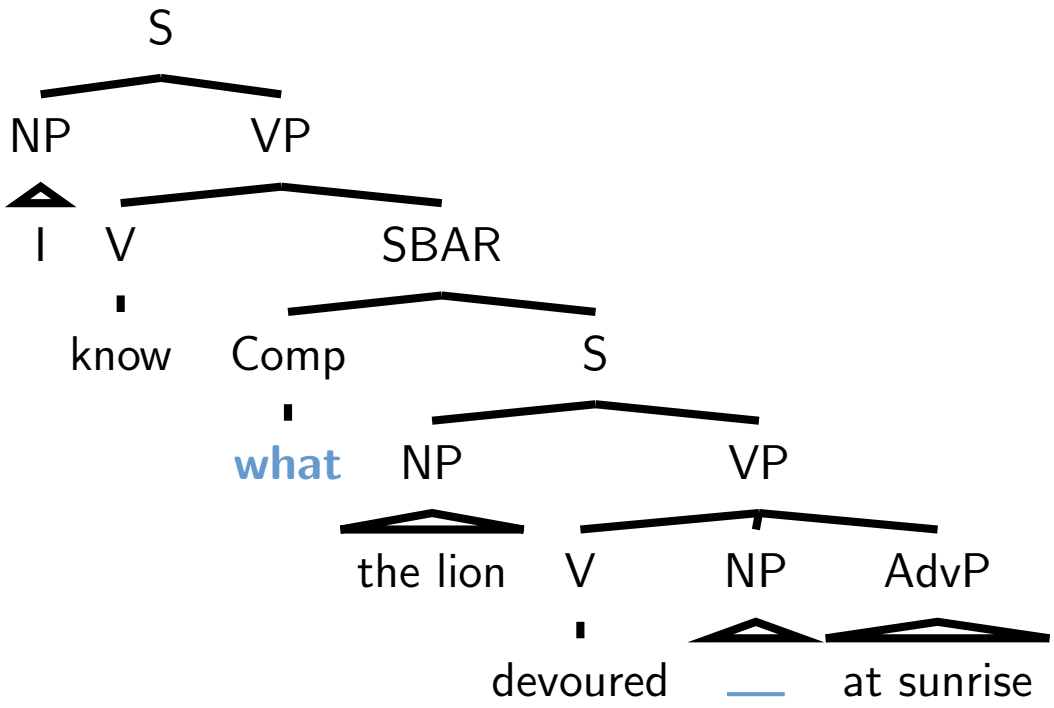


(Futrell, Wilcox, Morita, Qian, Ballesteros, & Levy, NAACL 2019)

✓ *I know what the lion devoured ___ at sunrise.*

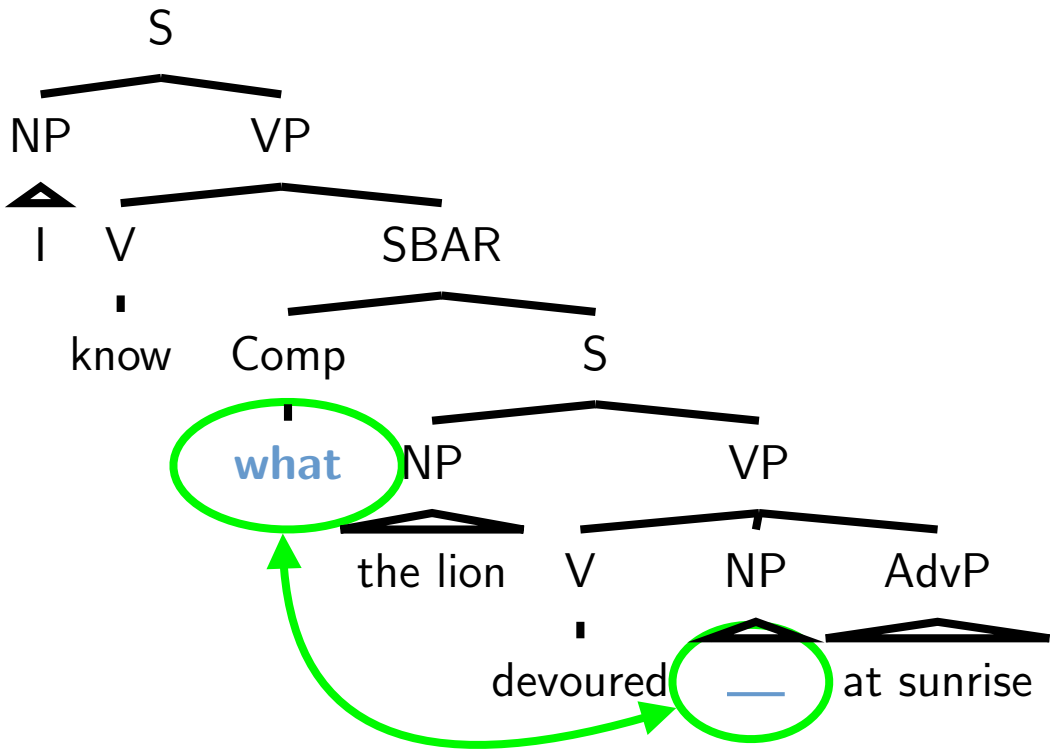
(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*



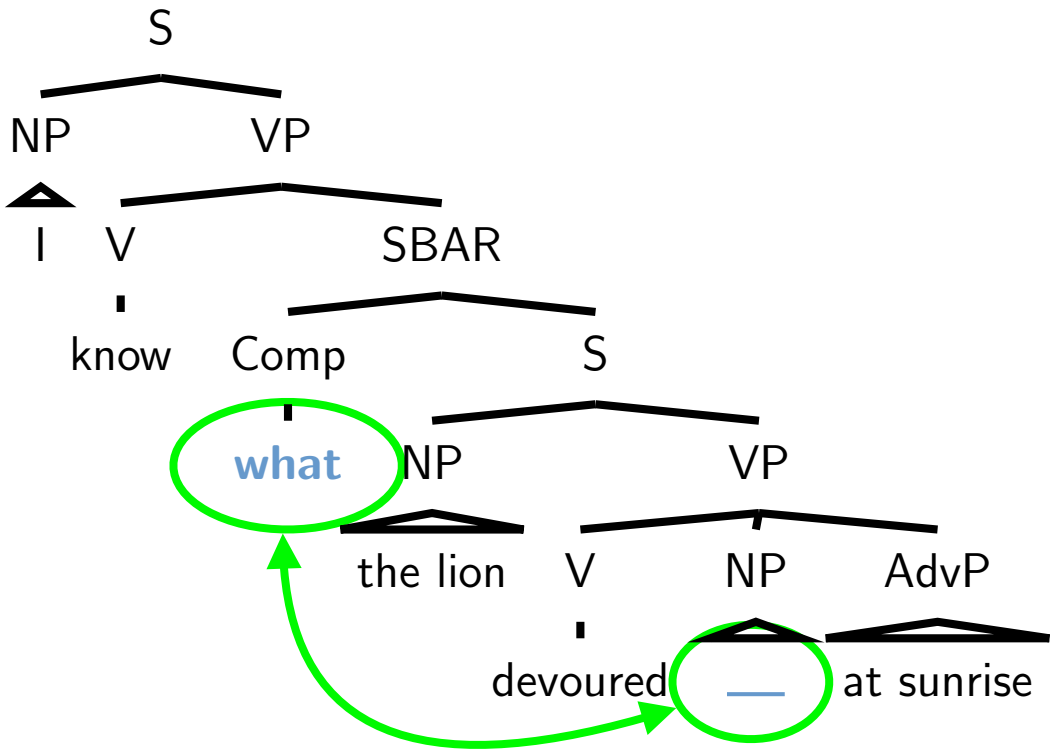
(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*



(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*

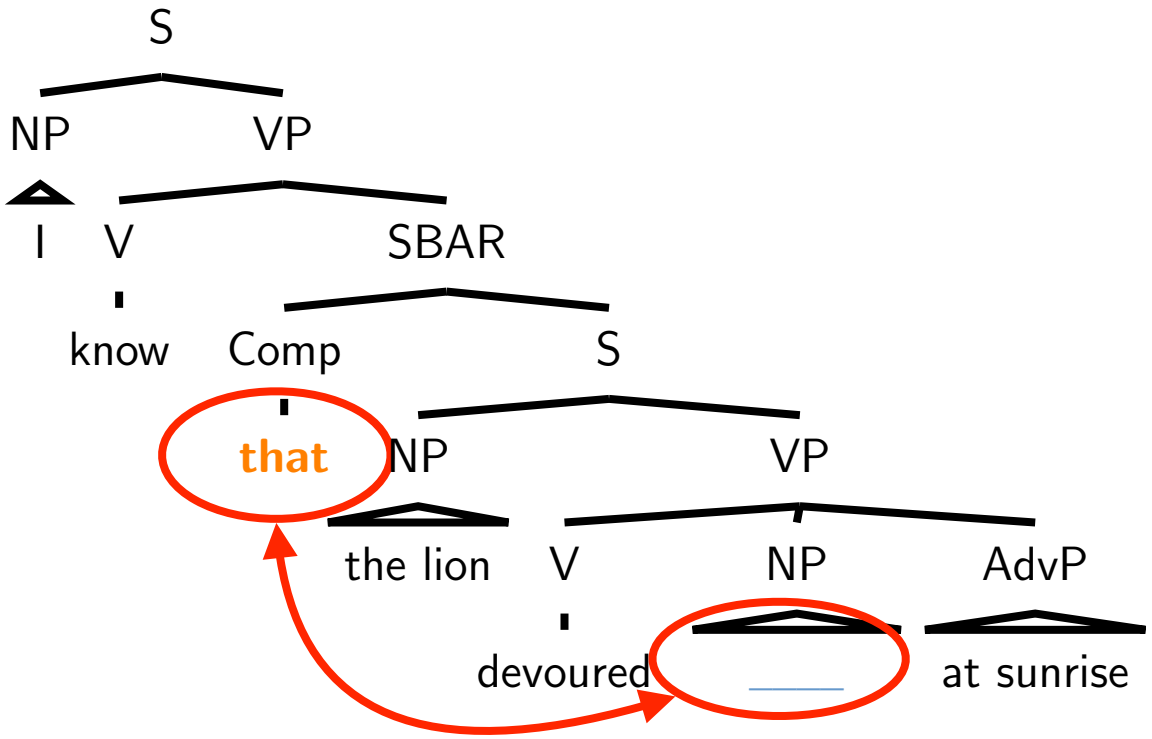


+FILLER
+GAP

(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*

∗ *I know that the lion devoured ___ at sunrise.*



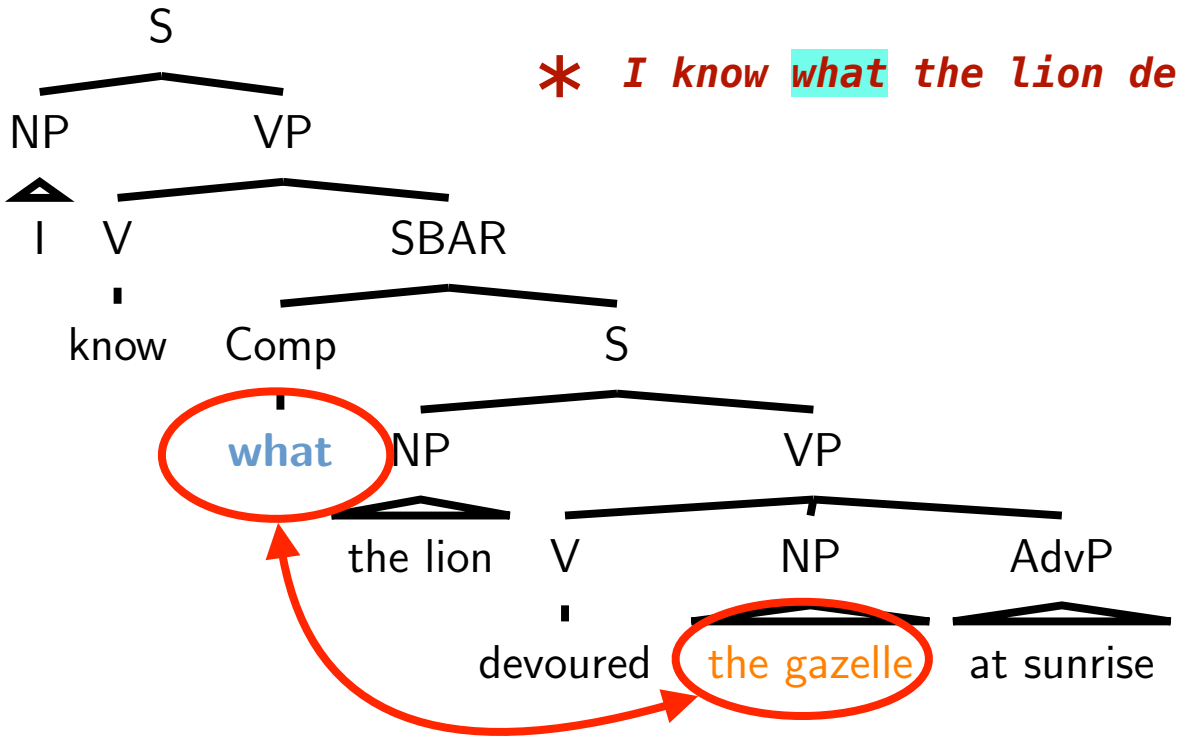
–FILLER
+GAP

(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*

∗ *I know that the lion devoured ___ at sunrise.*

∗ *I know what the lion devoured the gazelle at sunrise.*



+FILLER
–GAP

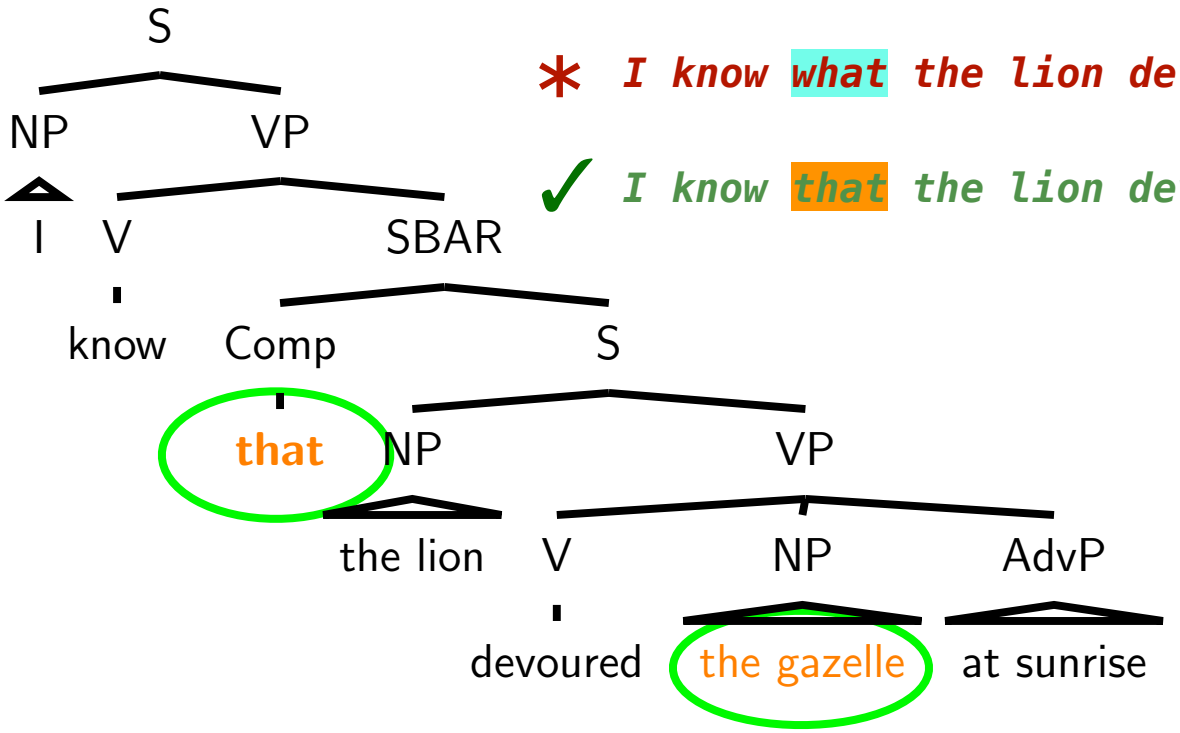
(Wilcox et al., 2018; in prep)

✓ *I know what the lion devoured ___ at sunrise.*

✱ *I know that the lion devoured ___ at sunrise.*

✱ *I know what the lion devoured the gazelle at sunrise.*

✓ *I know that the lion devoured the gazelle at sunrise.*



–FILLER

–GAP

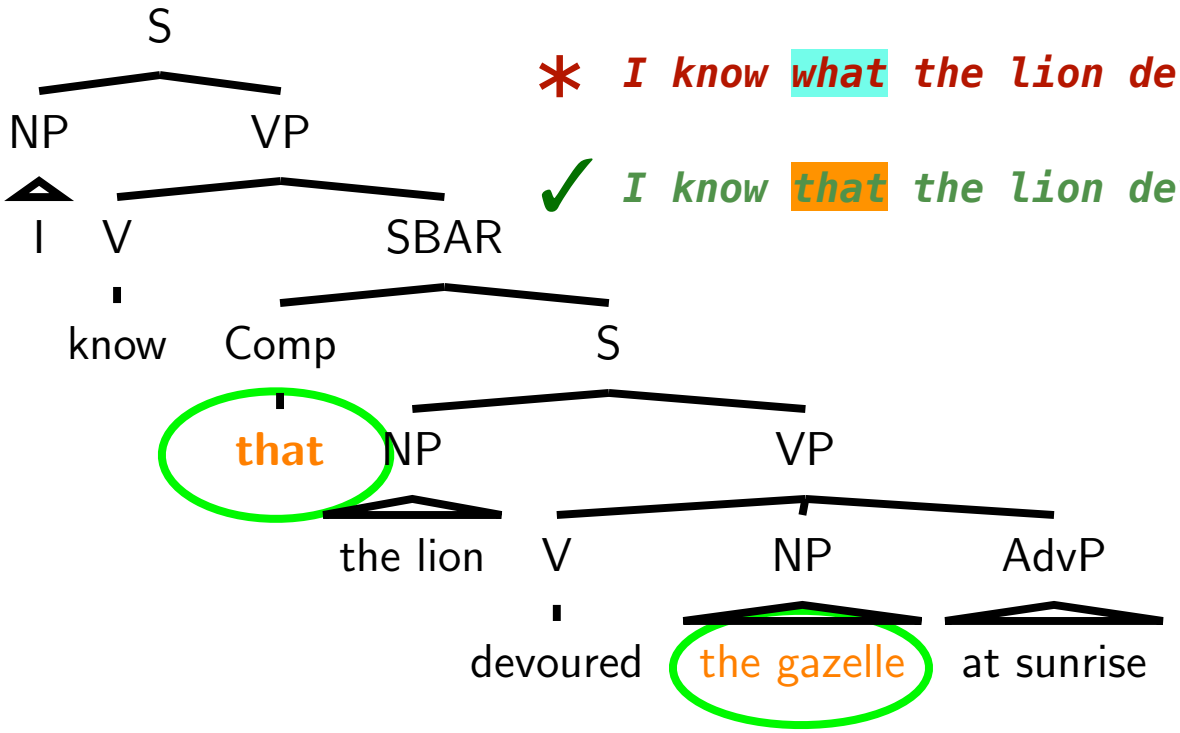
(Wilcox et al., 2018; in prep)

✓ I know what the lion devoured ___ at sunrise.

✱ I know that the lion devoured ___ at sunrise.

✱ I know what the lion devoured the gazelle at sunrise.

✓ I know that the lion devoured the gazelle at sunrise.



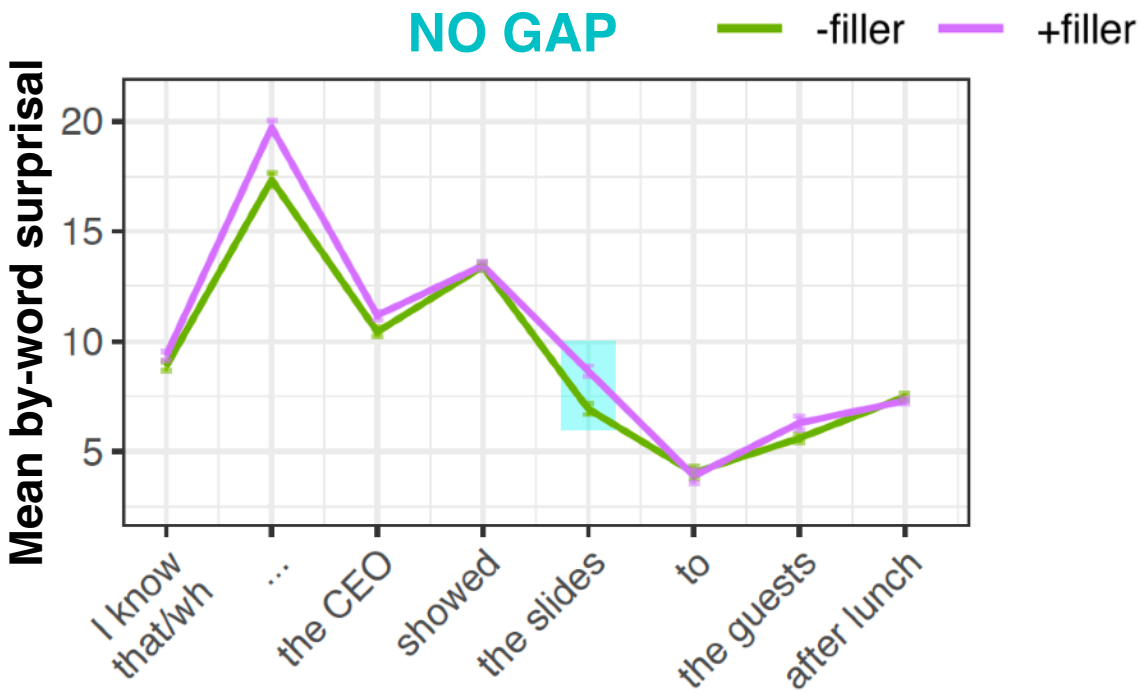
–FILLER
–GAP

Filler–gap dependencies *are a signature, theoretically central feature of natural language grammar*

(Wilcox et al., 2018; in prep)

✓ *I know that ... the CEO showed the slides to the guests after lunch.*

∗ *I know what ... the CEO showed the slides to the guests after lunch.*



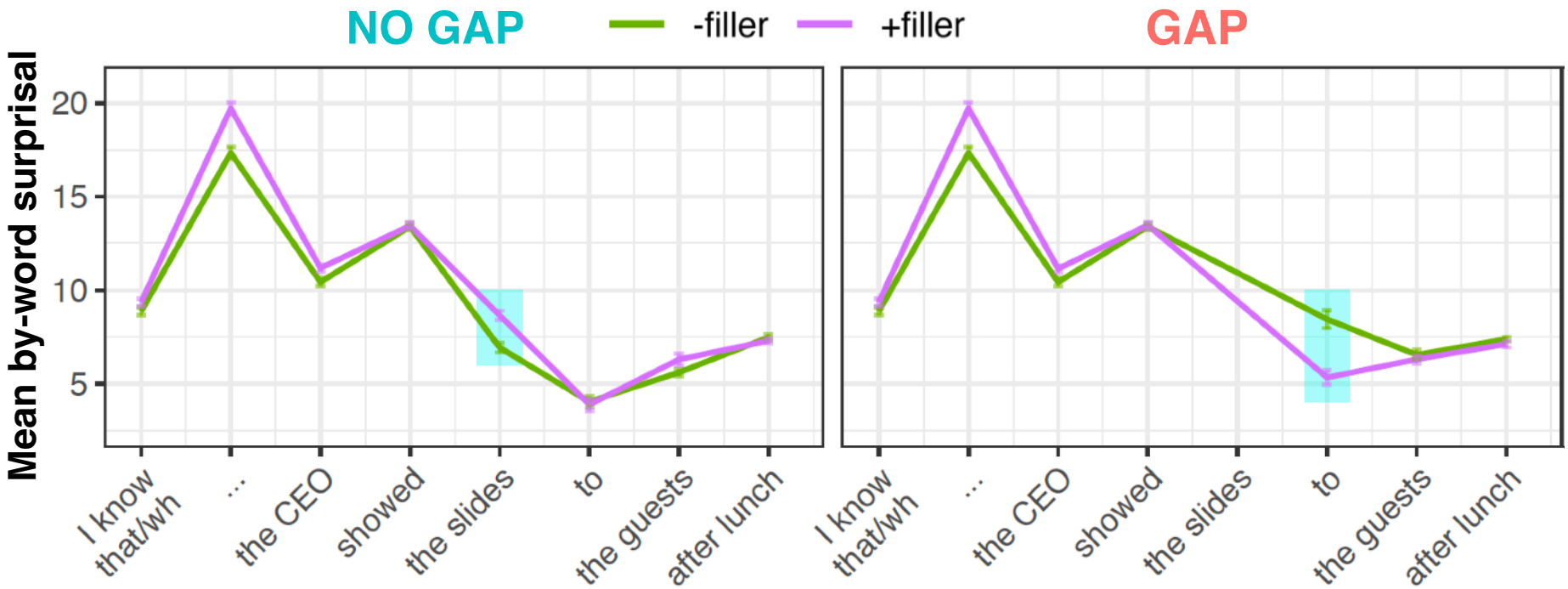
(Matches well-known human reading-time patterns: Stowe, 1986)

✓ I know that ... the CEO showed the slides to the guests after lunch.

∗ I know what ... the CEO showed the slides to the guests after lunch.

∗ I know that ... the CEO showed _____ to the guests after lunch.

✓ I know what ... the CEO showed _____ to the guests after lunch.



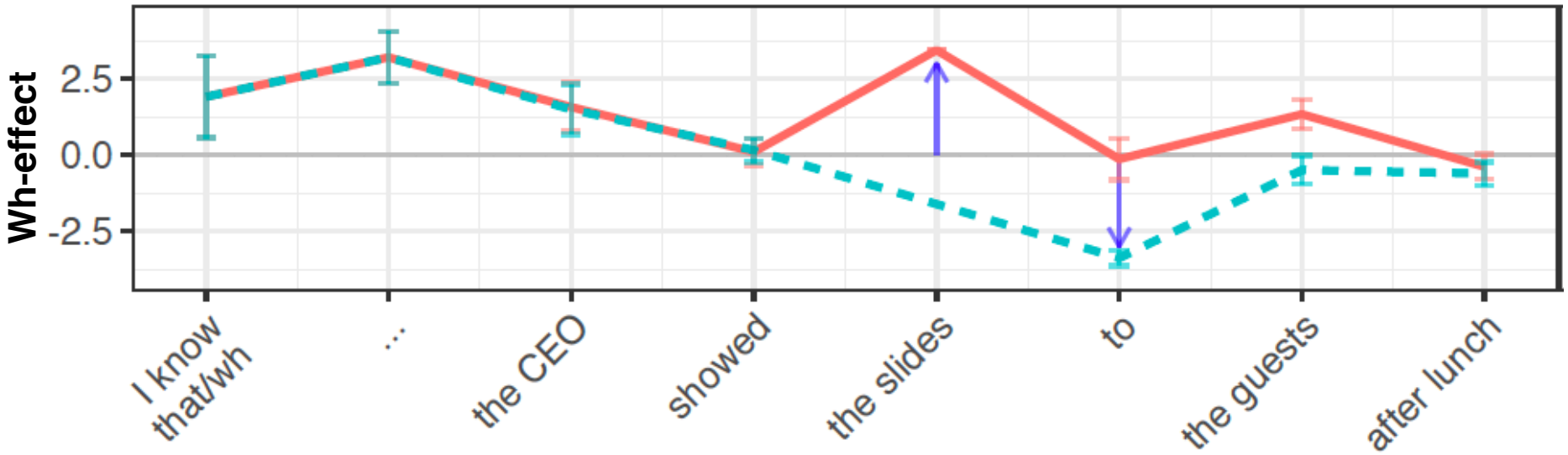
(Matches well-known human reading-time patterns: Stowe, 1986)

✓ I know that ... the CEO showed the slides to the guests after lunch.

∗ I know what ... the CEO showed the slides to the guests after lunch.

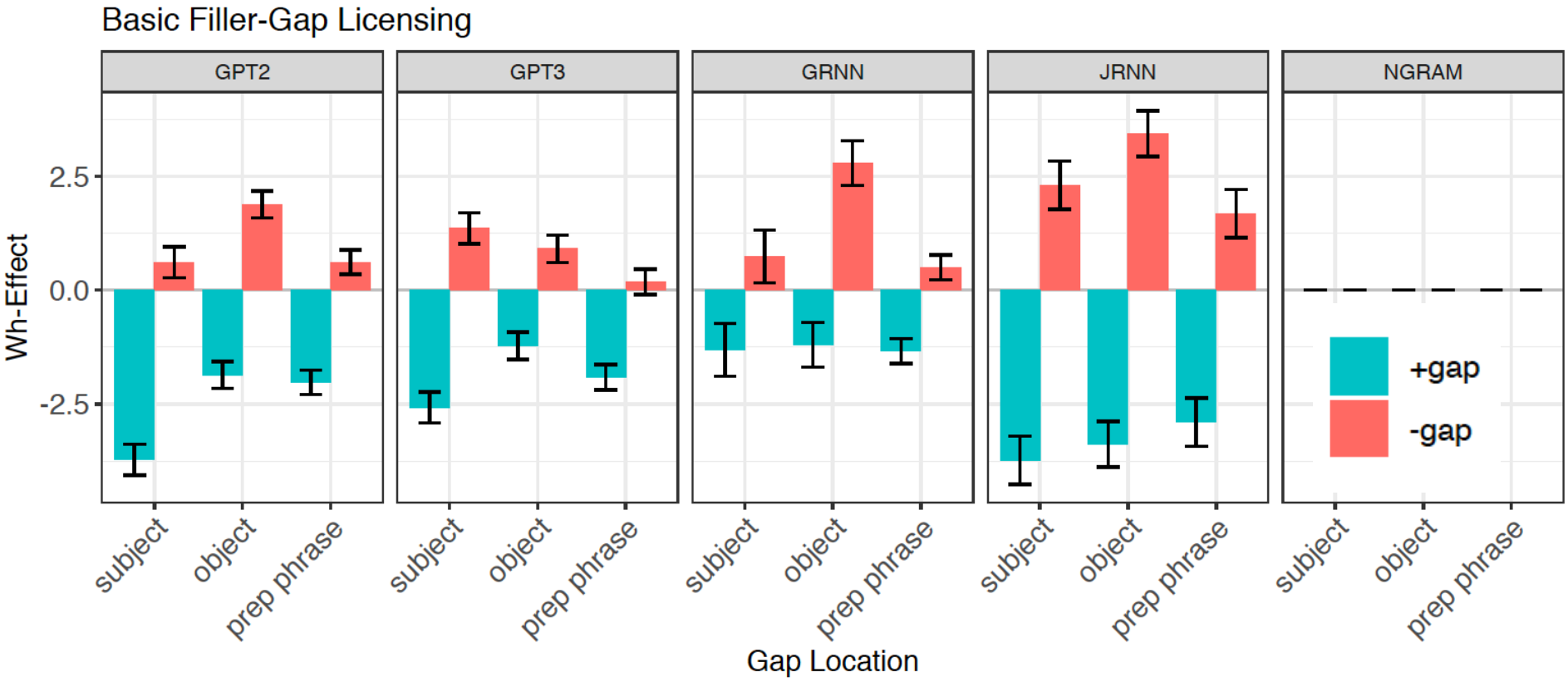
∗ I know that ... the CEO showed _____ to the guests after lunch.

✓ I know what ... the CEO showed _____ to the guests after lunch.



(Matches well-known human reading-time patterns: Stowe, 1986)

Flexibility of filler–gap dependencies



(Wilcox et al., 2018, in prep)

Unboundedness of *wh*-dependencies

I know what our mother gave __ to Mary last weekend.

Unboundedness of *wh*-dependencies

I know what our mother gave __ to Mary last weekend.

I know what our mother said that your friend gave __ to Mary last weekend.

Unboundedness of *wh*-dependencies

0 *I know what our mother gave __ to Mary last weekend.*

1 *I know what our mother said that your friend gave __ to Mary last weekend.*

Unboundedness of *wh*-dependencies

0 *I know what our mother gave __ to Mary last weekend.*

1 *I know what our mother said that your friend gave __ to Mary last weekend.*

2 *I know what our mother said that her friend remarked that your friend gave __ to Mary last weekend.*

Unboundedness of *wh*-dependencies

0 *I know what our mother gave __ to Mary last weekend.*

1 *I know what our mother said that your friend gave __ to Mary last weekend.*

2 *I know what our mother said that her friend remarked that your friend gave __ to Mary last weekend.*

3 *I know what our mother said that her friend remarked that the park attendant wondered that your friend gave __ to Mary last weekend.*

Unboundedness of *wh*-dependencies

0 *I know what our mother gave __ to Mary last weekend.*

1 *I know what our mother said that your friend gave __ to Mary last weekend.*

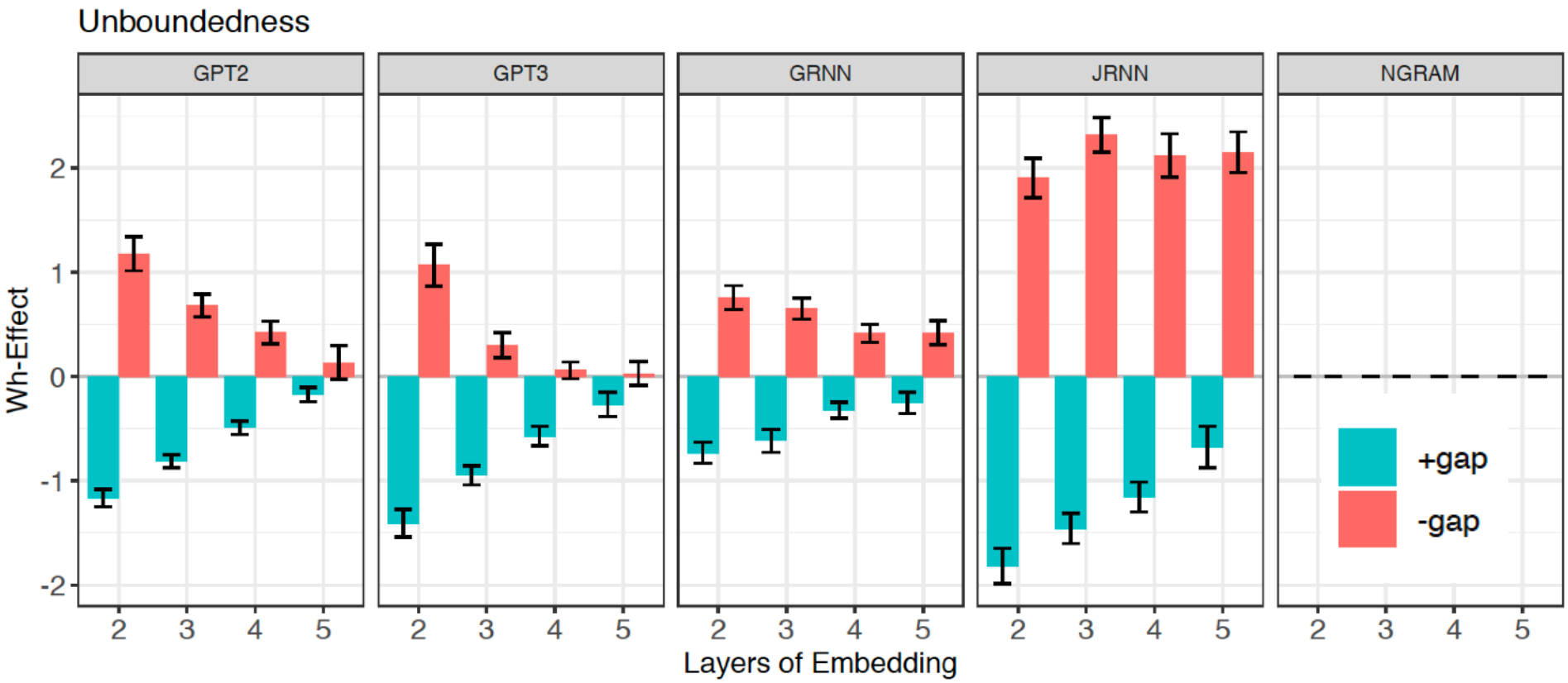
2 *I know what our mother said that her friend remarked that your friend gave __ to Mary last weekend.*

3 *I know what our mother said that her friend remarked that the park attendant wondered that your friend gave __ to Mary last weekend.*

4 *I know what our mother said that her friend remarked that the park attendant wondered that the people stated that your friend gave __ to Mary last weekend.*

Unboundedness of filler–gap dependency

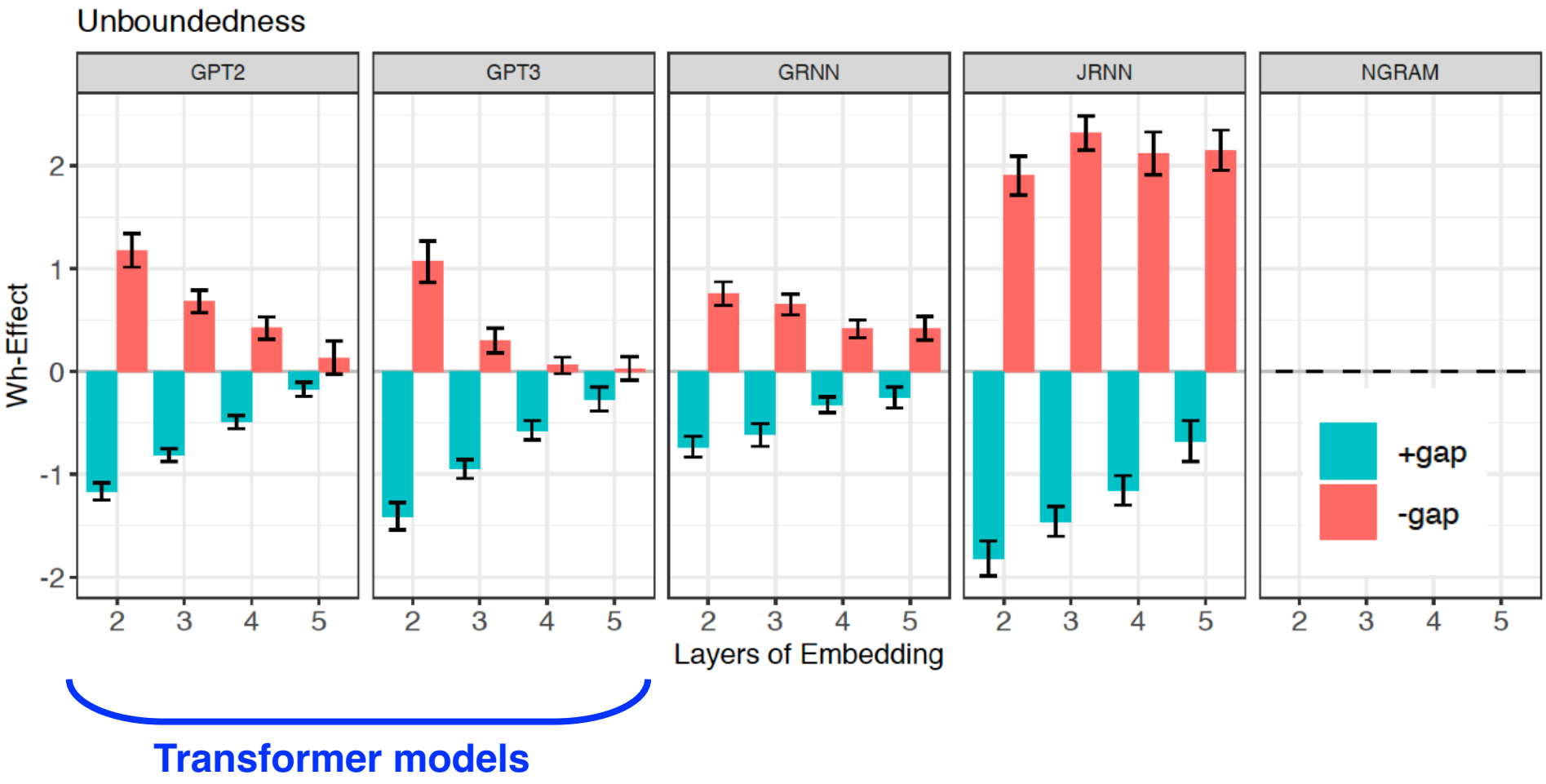
- For object gaps:



(Wilcox et al., in prep)

Unboundedness of filler–gap dependency

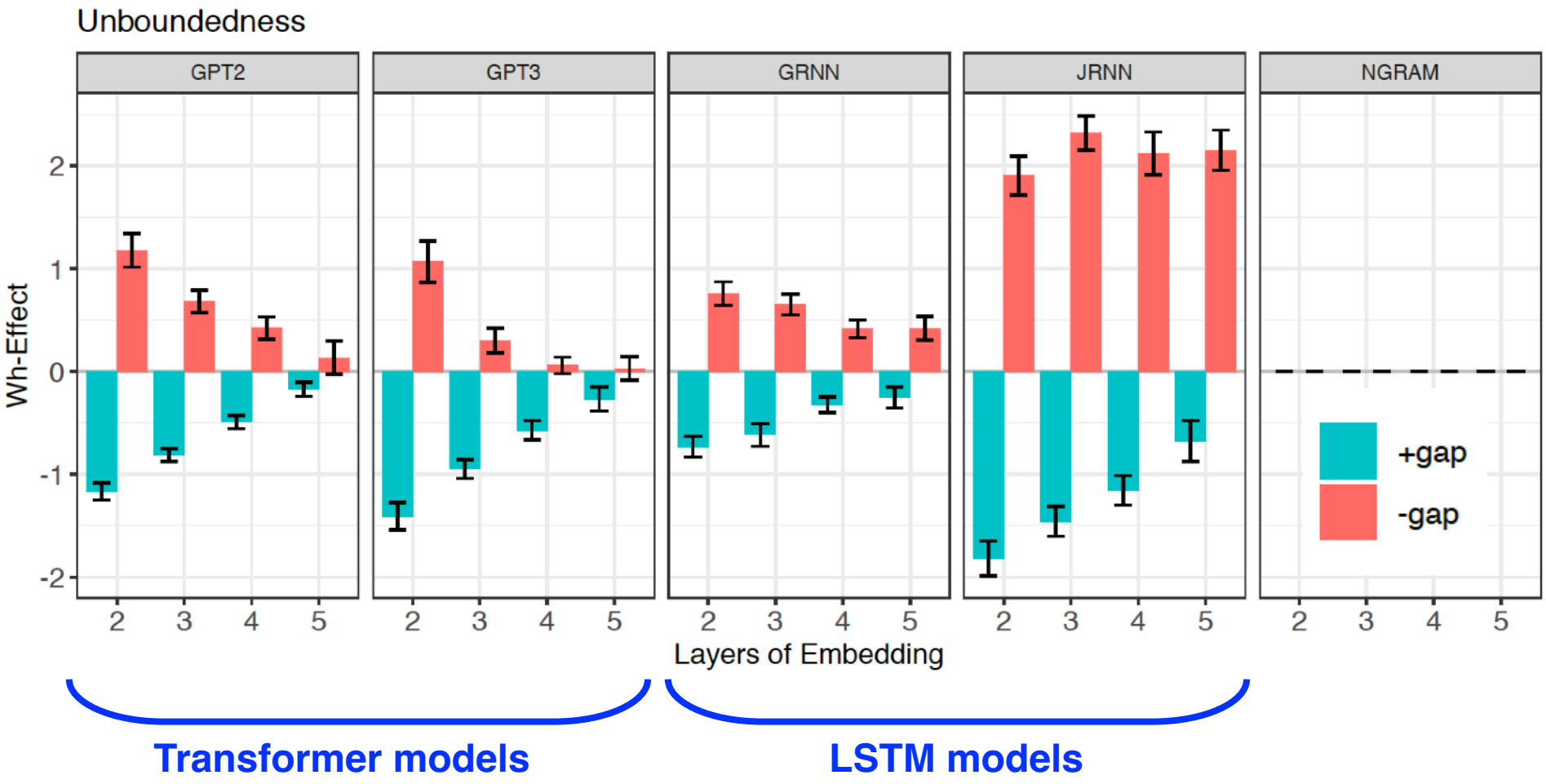
- For object gaps:



(Wilcox et al., in prep)

Unboundedness of filler–gap dependency

- For object gaps:



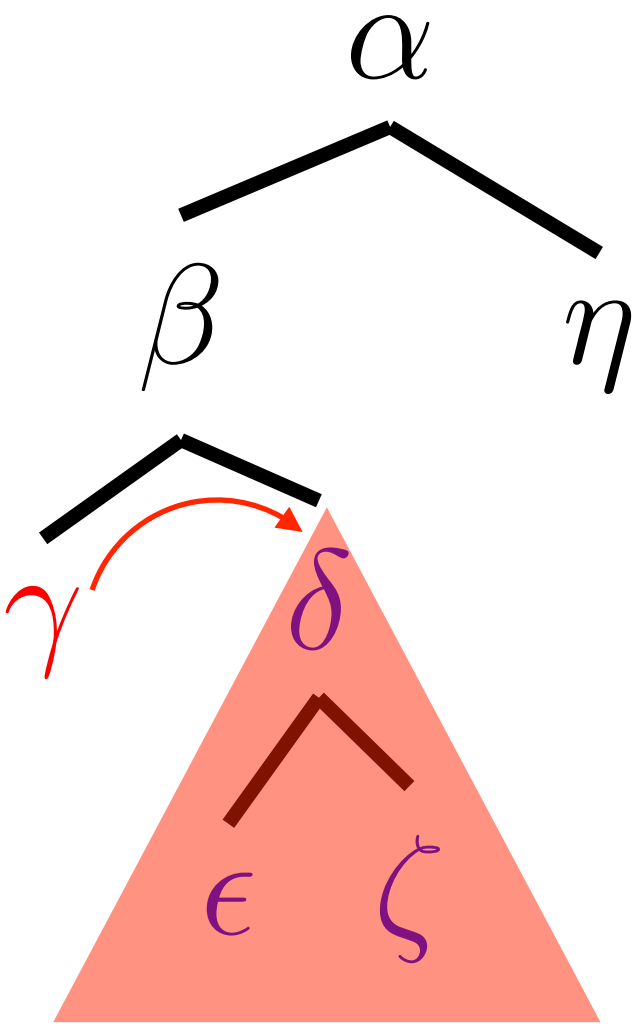
(Wilcox et al., in prep)

Potential concern #1

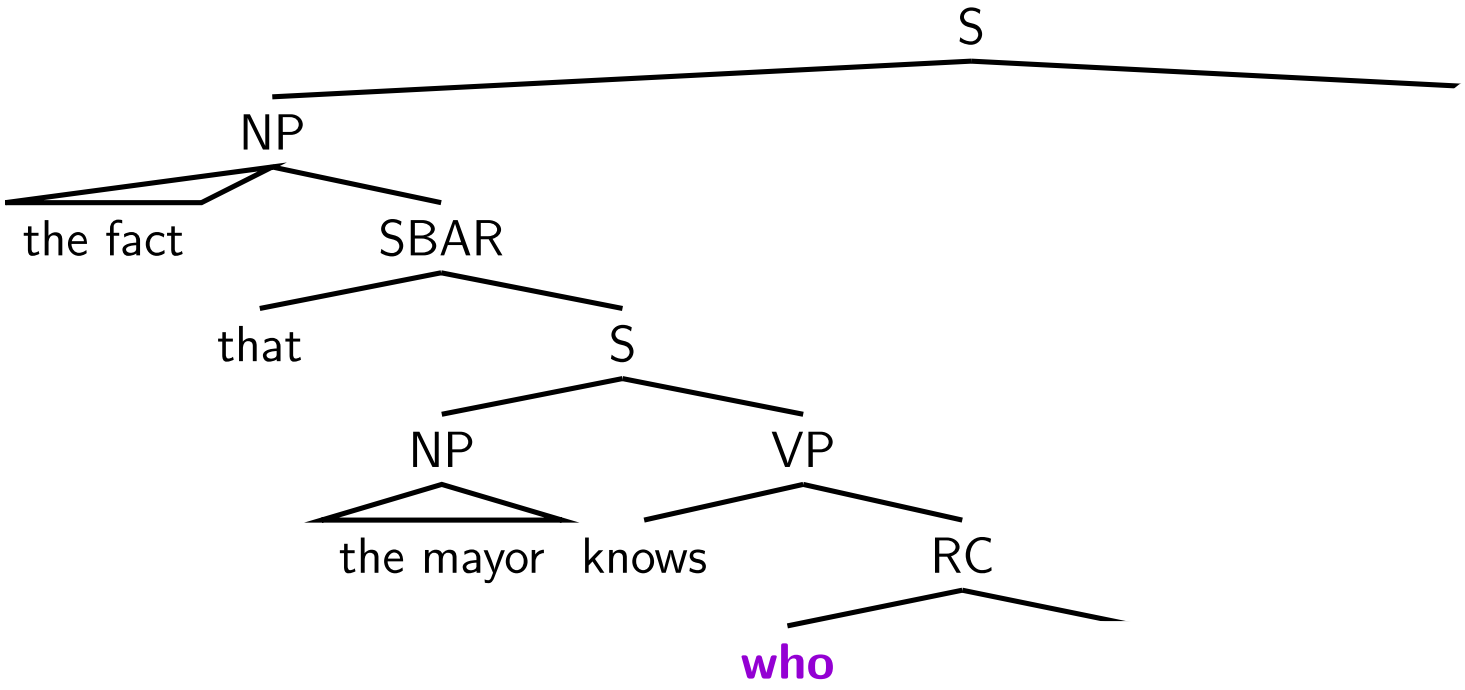
Couldn't the models be learning a ***linear*** dependency between filler and gap, not a ***hierarchical*** dependency?

Syntactic Hierarchy

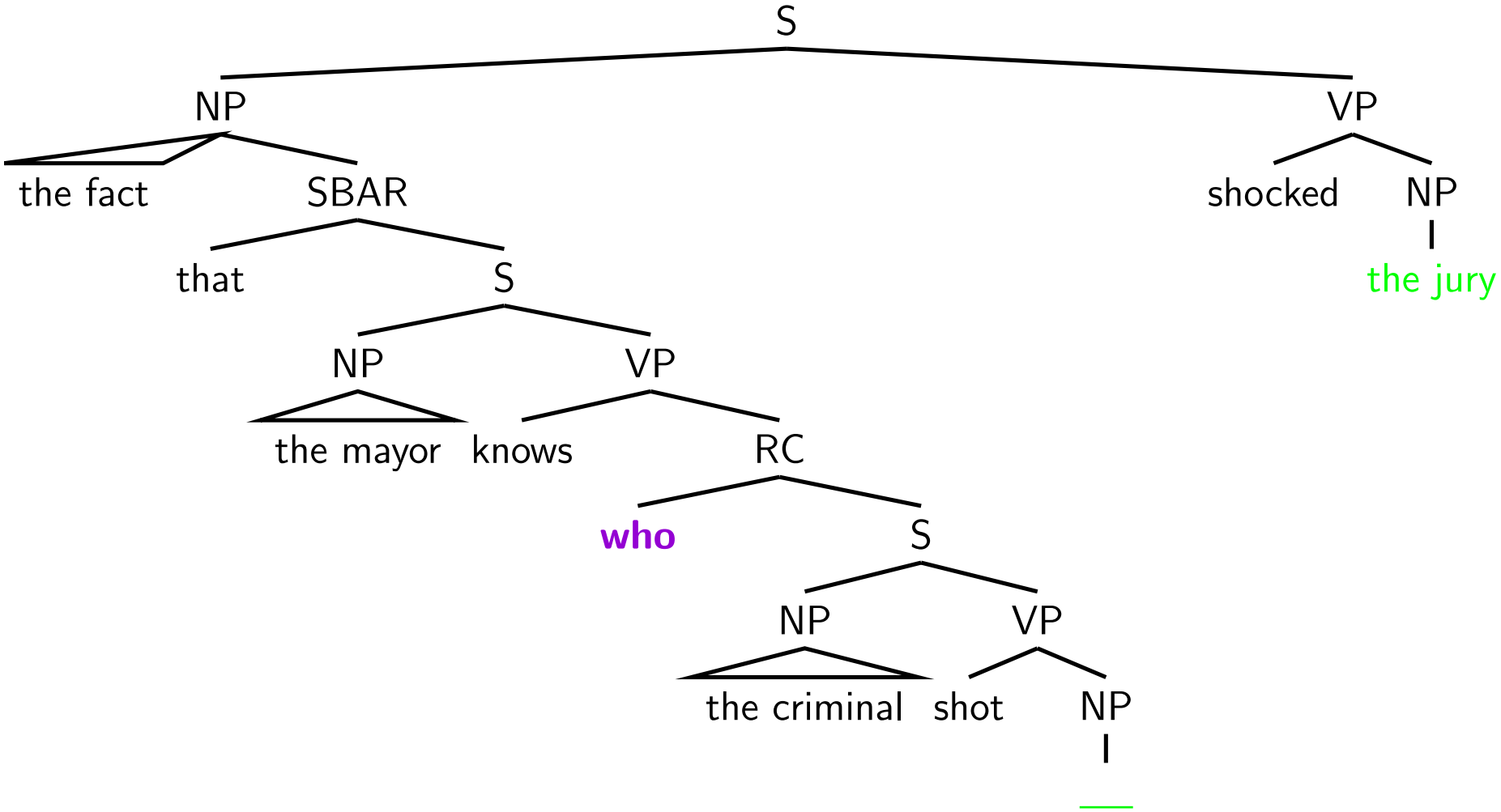
- A filler must be appropriately "above" its gap



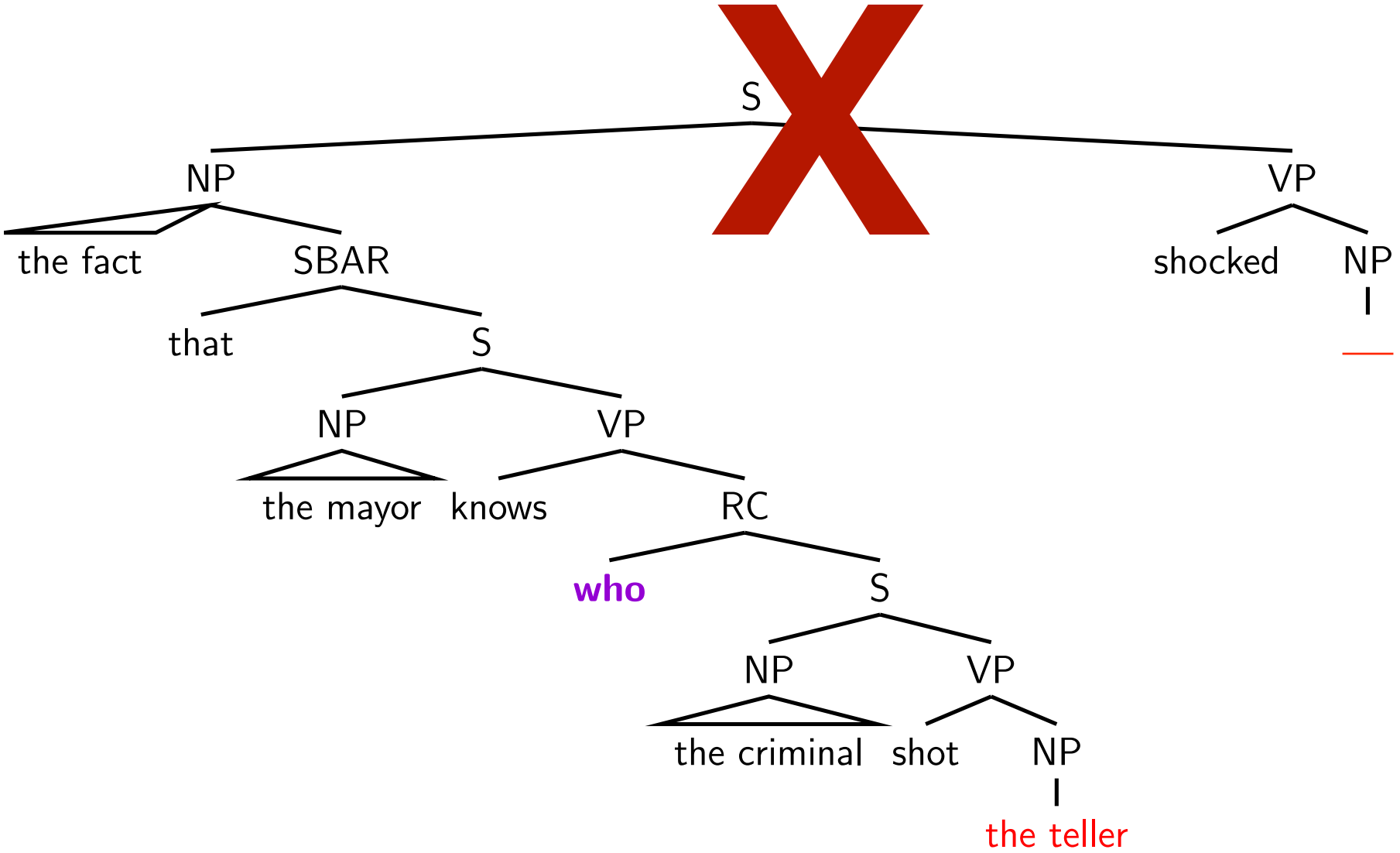
(Wilcox et al., 2019; in prep)



(Wilcox et al., 2019; in prep)



(Wilcox et al., 2019; in prep)



(Wilcox et al., 2019; in prep)

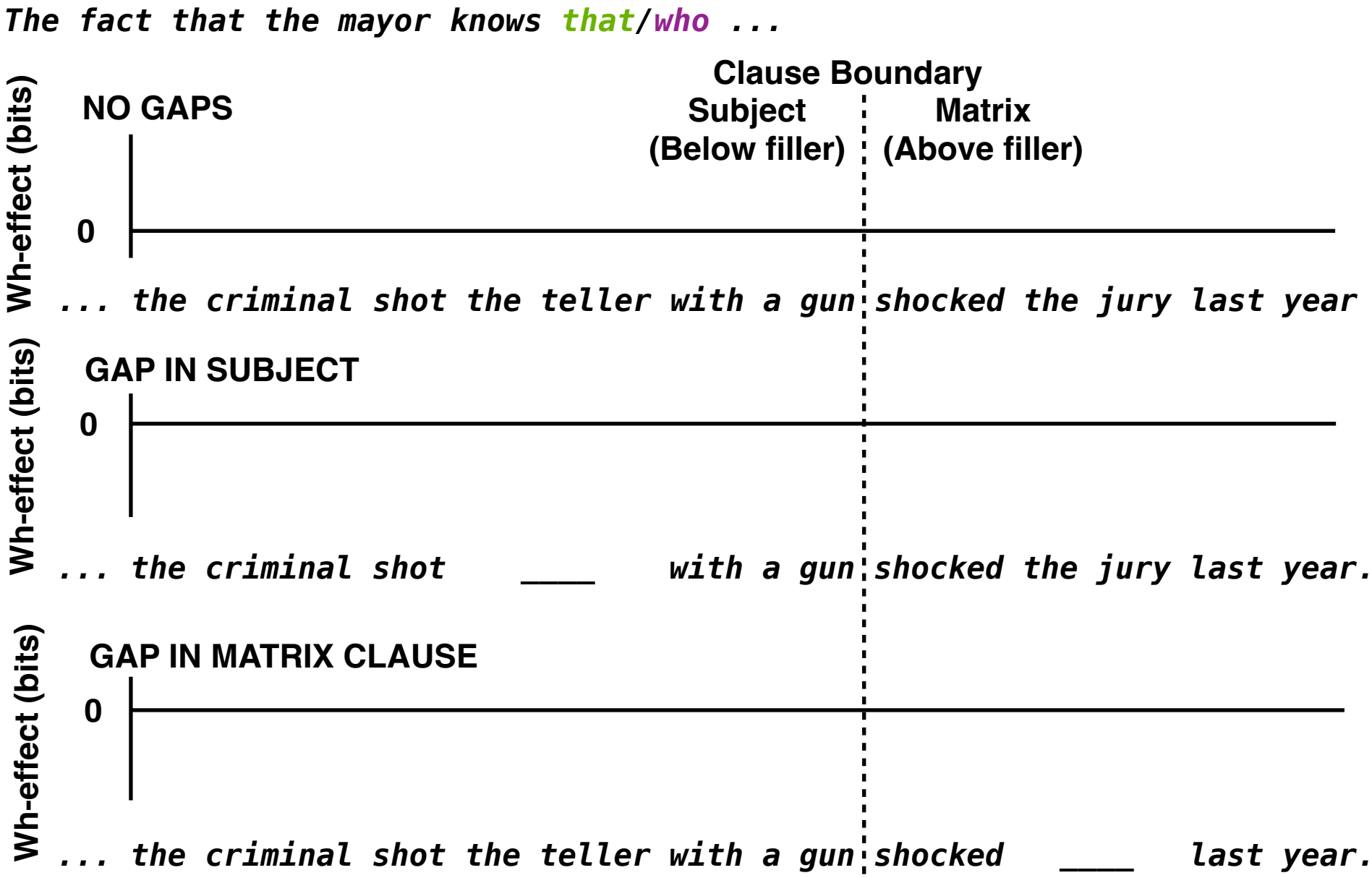
Predictions for human-like processing

The fact that the mayor knows that/who ...

	Clause Boundary: Subject (Below filler)	Clause Boundary: Matrix (Above filler)
NO GAPS		
	... the criminal shot the teller with a gun	*shocked the jury last year*
GAP IN SUBJECT		
	... the criminal shot ____ with a gun	*shocked the jury last year.*
GAP IN MATRIX CLAUSE		
	... the criminal shot the teller with a gun	*shocked ____ last year.*

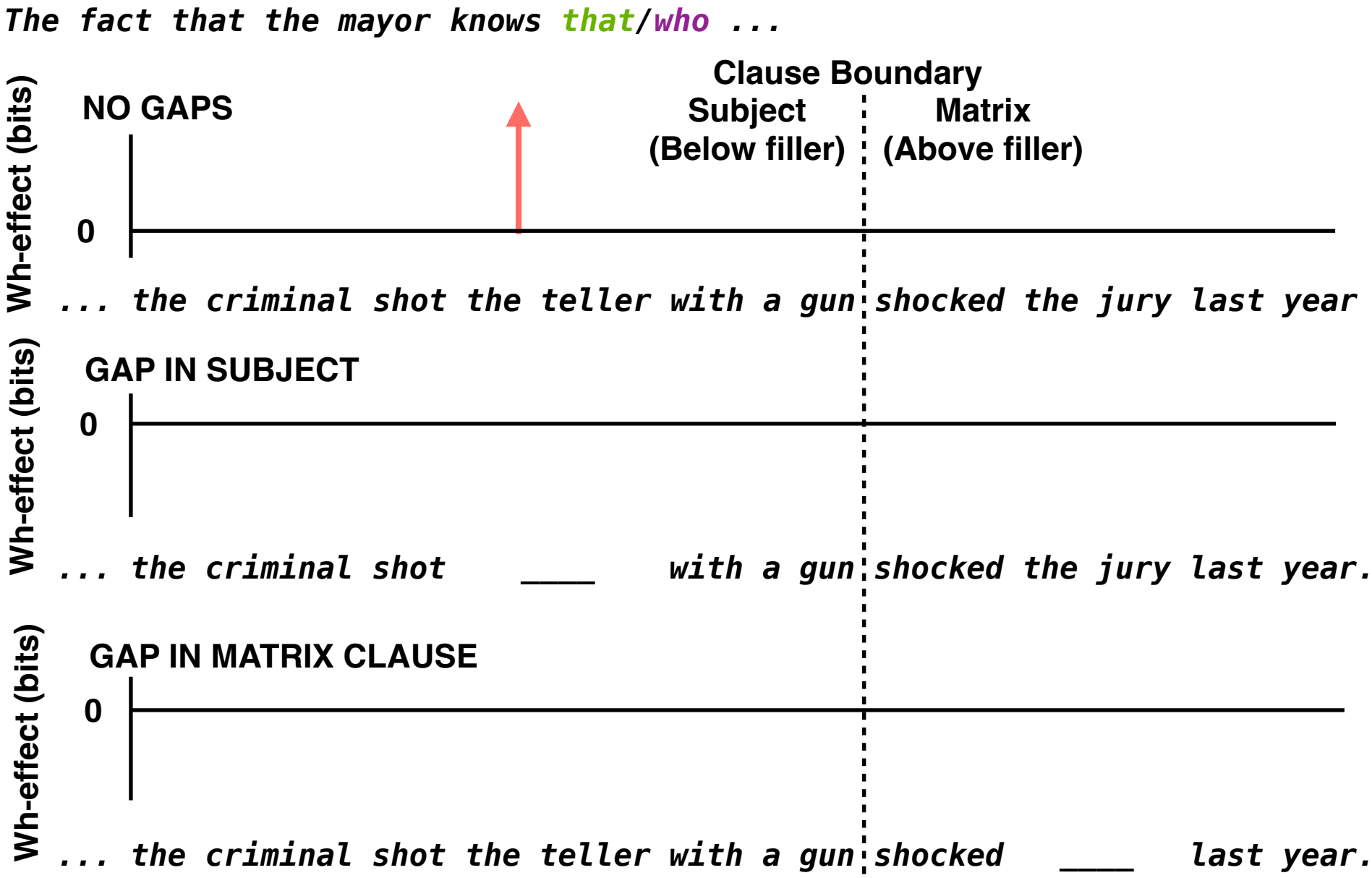
(Wilcox et al., 2019; in prep)

Predictions for human-like processing



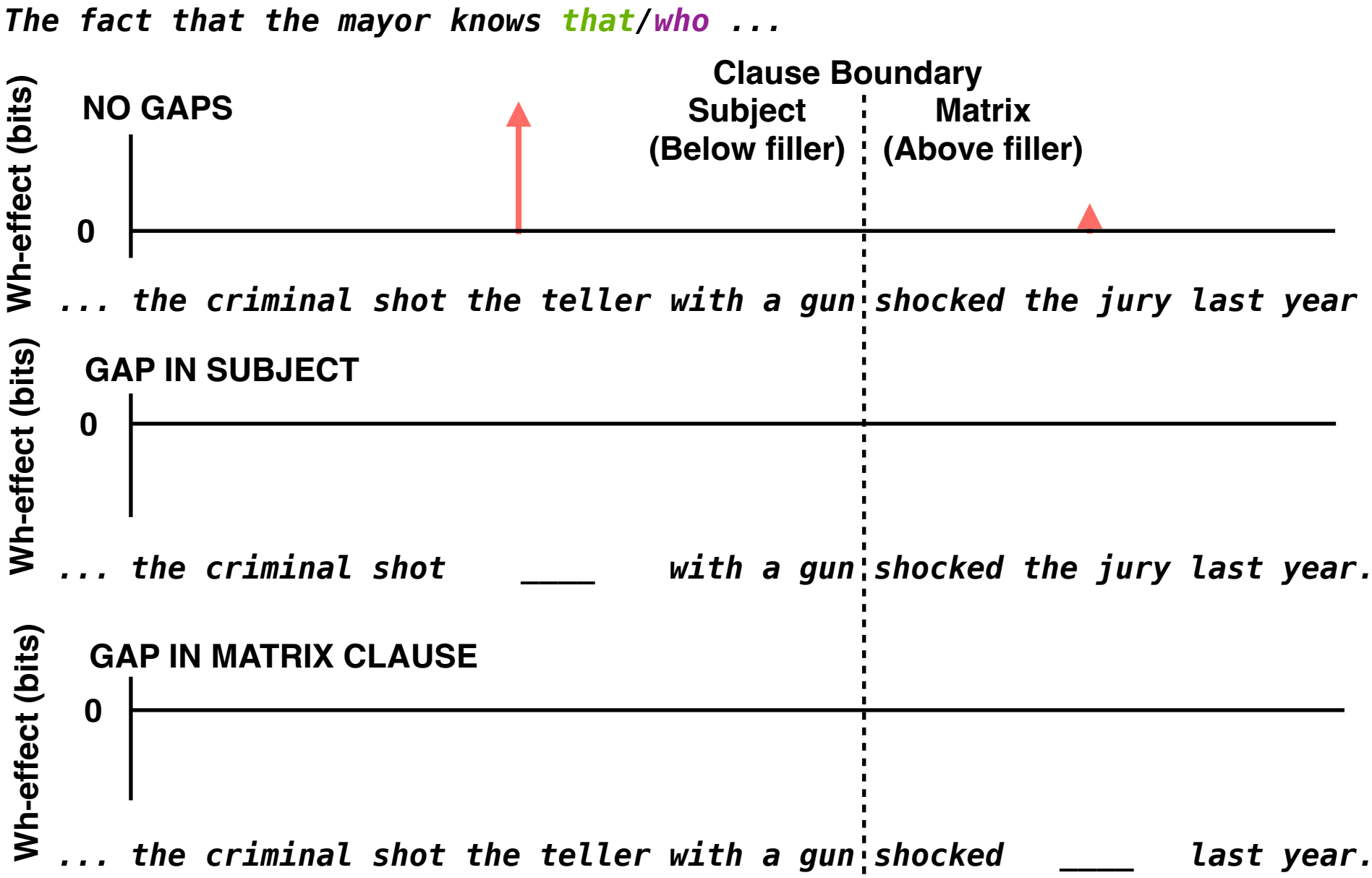
(Wilcox et al., 2019; in prep)

Predictions for human-like processing



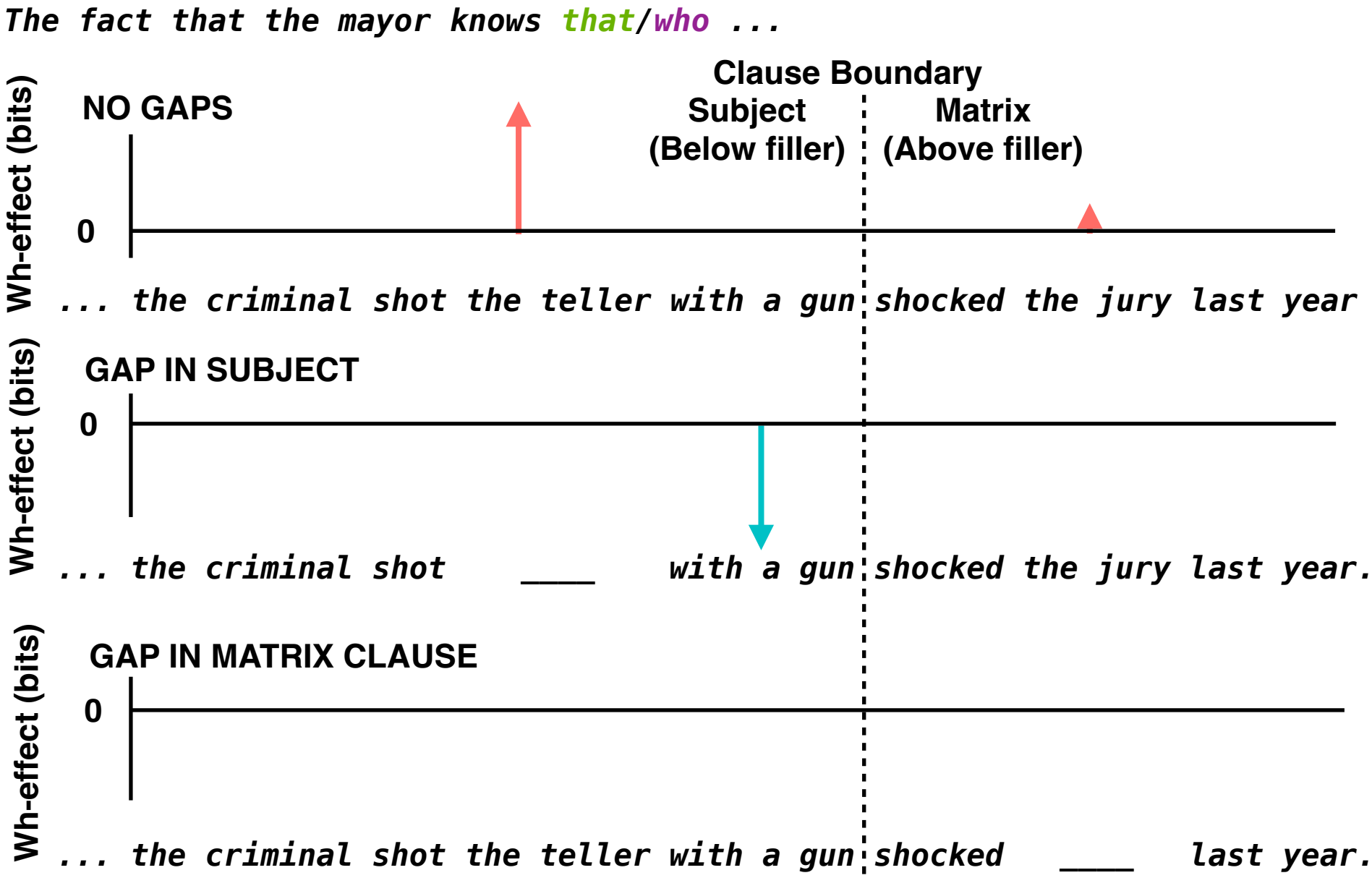
(Wilcox et al., 2019; in prep)

Predictions for human-like processing



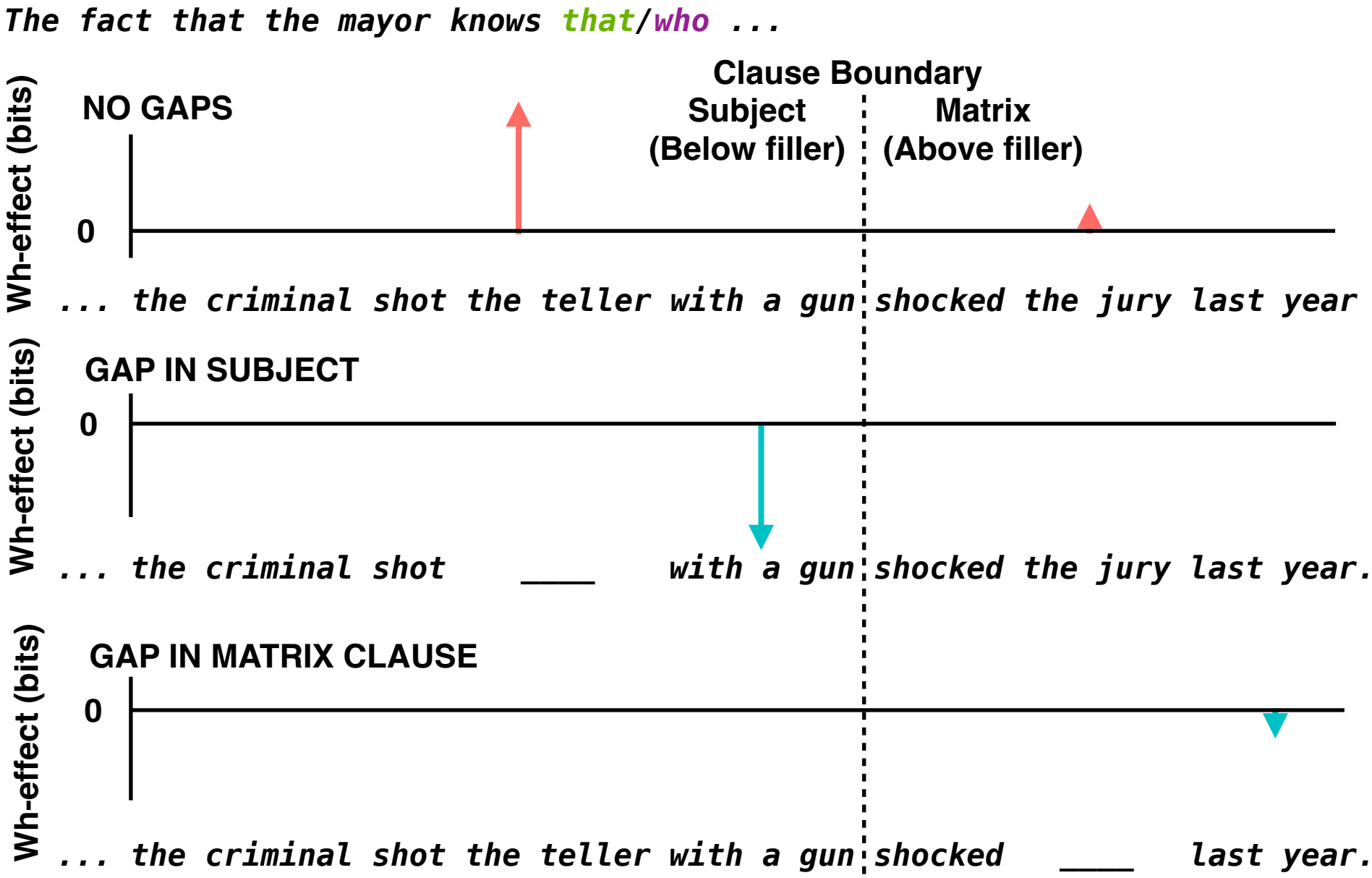
(Wilcox et al., 2019; in prep)

Predictions for human-like processing



(Wilcox et al., 2019; in prep)

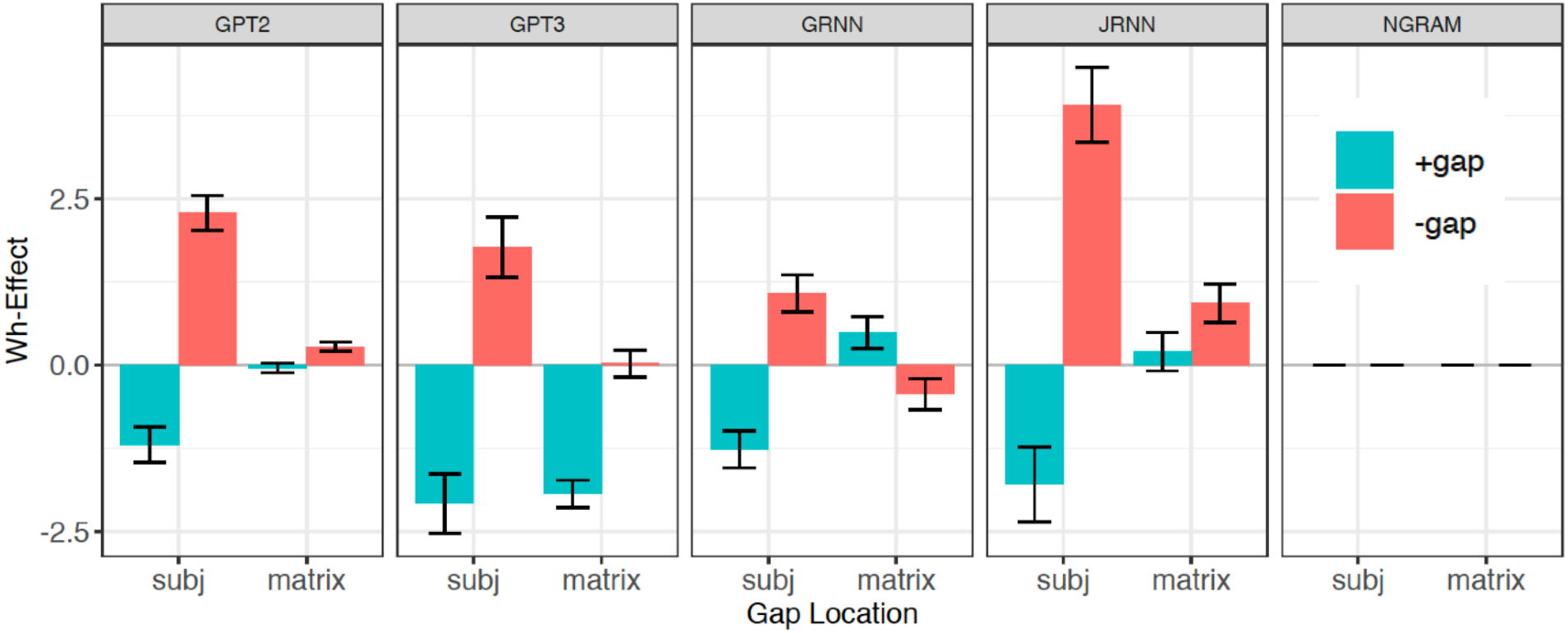
Predictions for human-like processing



(Wilcox et al., 2019; in prep)

Sensitivity to syntactic hierarchy

- For object gaps:



(Wilcox et al., in prep)

Potential concern #1

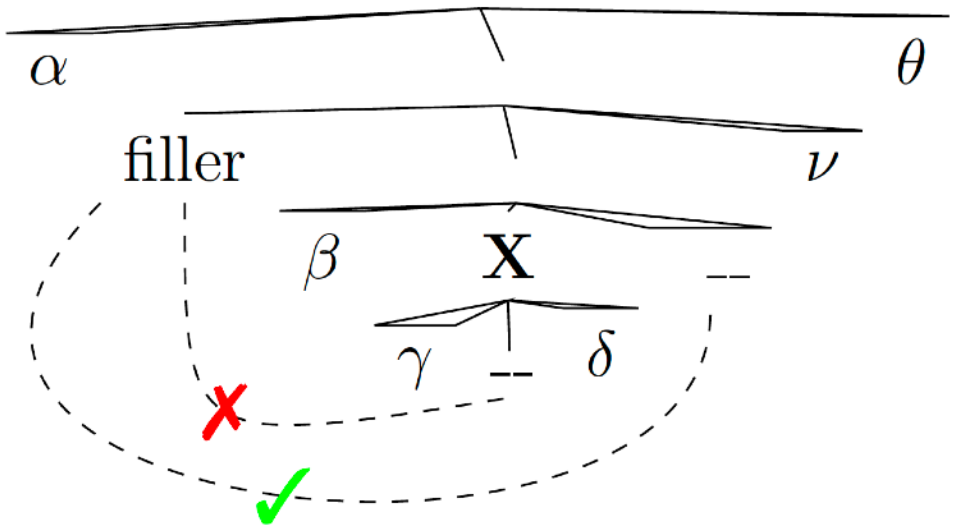
Couldn't the models be learning a ***linear*** dependency between filler and gap, not a ***hierarchical*** dependency?

Potential concern #1 — *addressed*

Couldn't the models be learning a *linear* dependency between filler and gap, not a *hierarchical* dependency?

Our results suggest that deep-learning models trained on enough data are sensitive to syntactic hierarchy for *wh*-dependency

Syntactic island constraints

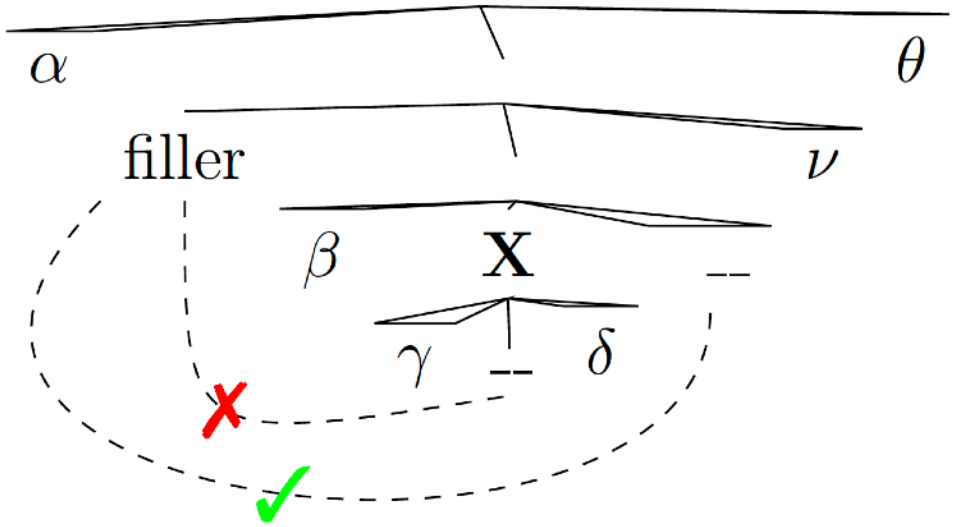


(Phillips, 2013; see also Pearl & Sprouse, 2013)

(Wilcox et al., 2018; in prep)

Syntactic island constraints

- Some types of phrases are ***islands***: filler–gap dependencies cannot link from outside to inside of them

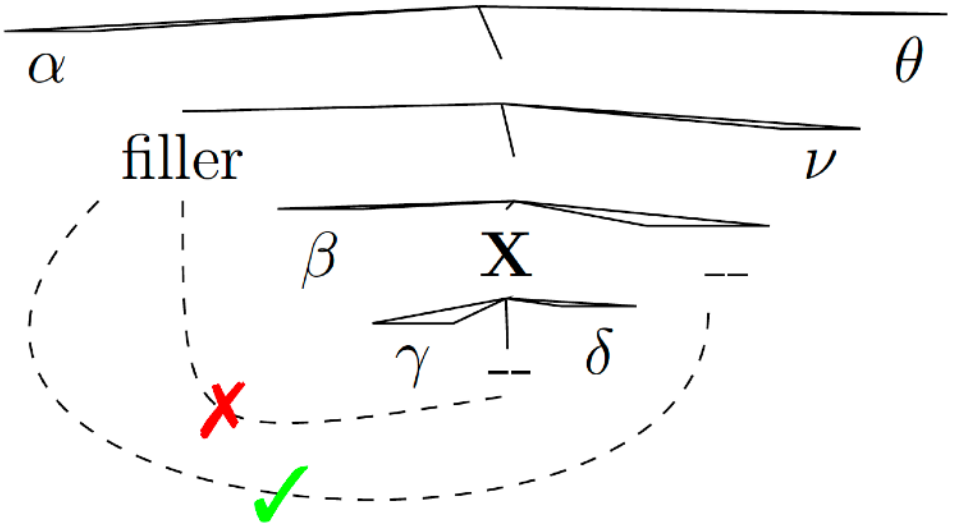


(Phillips, 2013; see also Pearl & Sprouse, 2013)

(Wilcox et al., 2018; in prep)

Syntactic island constraints

- Some types of phrases are ***islands***: filler–gap dependencies cannot link from outside to inside of them



- "Island effects have long been regarded as strong motivation for domain-specific innate constraints on human language...likely to be difficult to observe in the input that children must learn from."

(Phillips, 2013; see also Pearl & Sprouse, 2013)

(Wilcox et al., 2018; in prep)

Syntactic islands

***Wh*-complementizers** block filler—gap dependencies:

I know what Alex said…

✓ *…your friend devoured __ at the party.*
[no complementizer]

(Wilcox et al., 2018; in prep)

Syntactic islands

***Wh*-complementizers** block filler—gap dependencies:

I know what Alex said…

✓ *…your friend devoured __ at the party.*
[no complementizer]

(✓) ***…that*** *your friend devoured __ at the party.*
[*that* complementizer]

(Wilcox et al., 2018; in prep)

Syntactic islands

***Wh*-complementizers** block filler—gap dependencies:

I know what Alex said…

✓ *…your friend devoured __ at the party.*
[no complementizer]

(✓) ***…that** your friend devoured __ at the party.*
[*that* complementizer]

* ***…whether** your friend devoured __ at the party.*
[*wh*-complementizer]

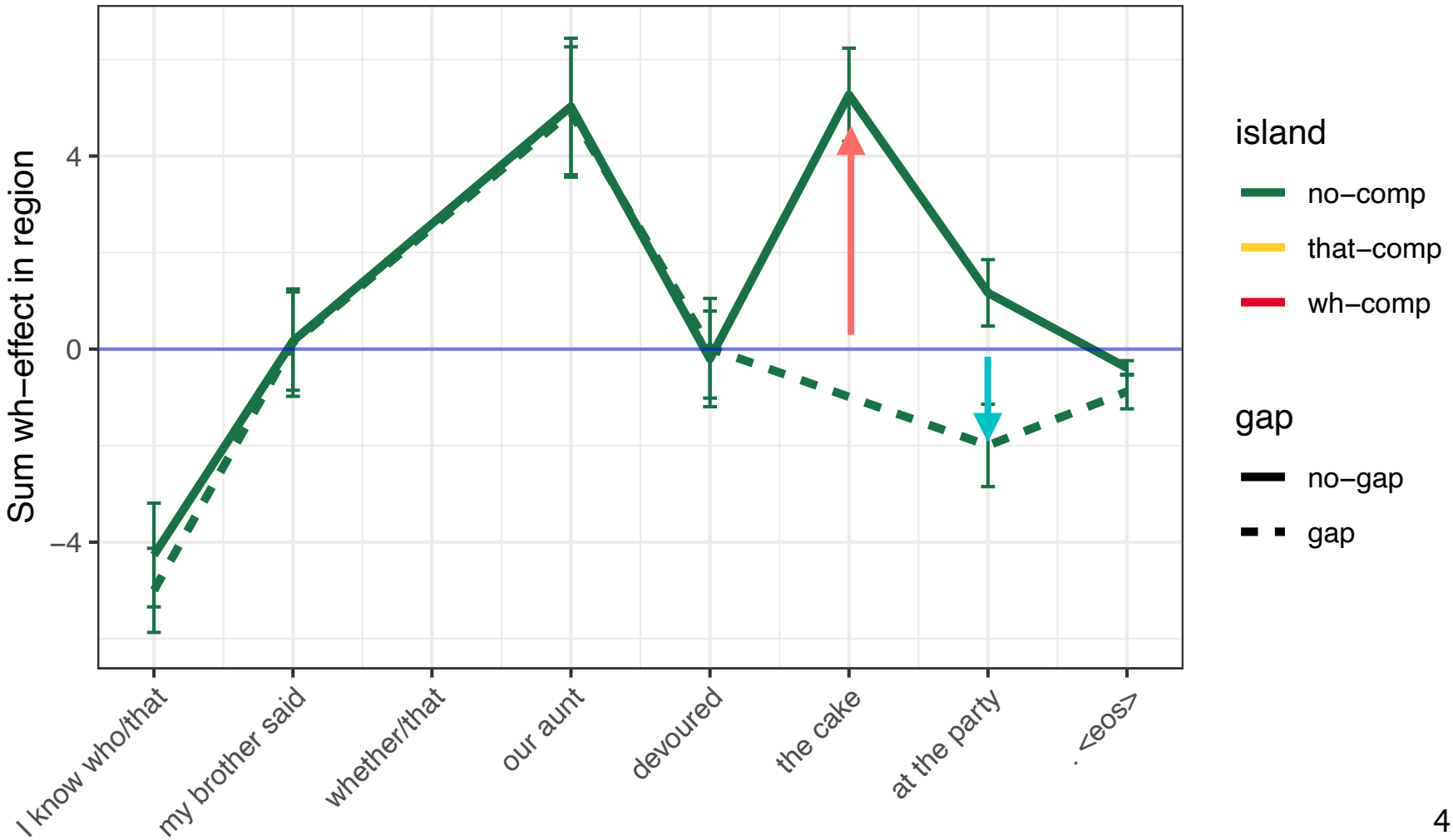
(Wilcox et al., 2018; in prep)

✓ *I know that my brother said our aunt devoured the cake at the party.*

* *I know what my brother said our aunt devoured the cake at the party.*

* *I know that my brother said our aunt devoured _____ at the party.*

✓ *I know what my brother said our aunt devoured _____ at the party.*

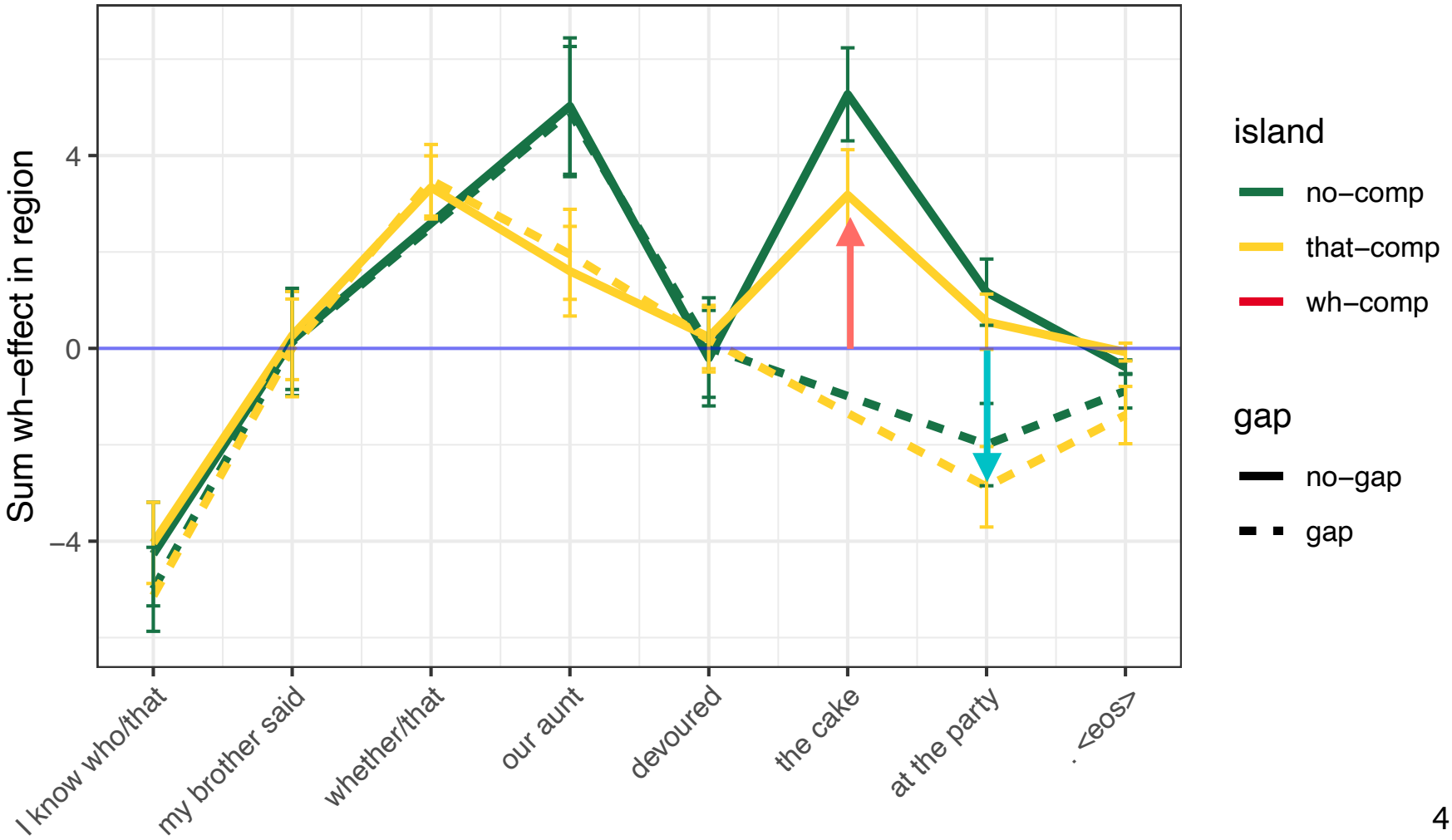


✓ *I know that my brother said that our aunt devoured the cake at the party.*

* *I know what my brother said that our aunt devoured the cake at the party.*

* *I know that my brother said that our aunt devoured _____ at the party.*

✓ *I know what my brother said that our aunt devoured _____ at the party.*

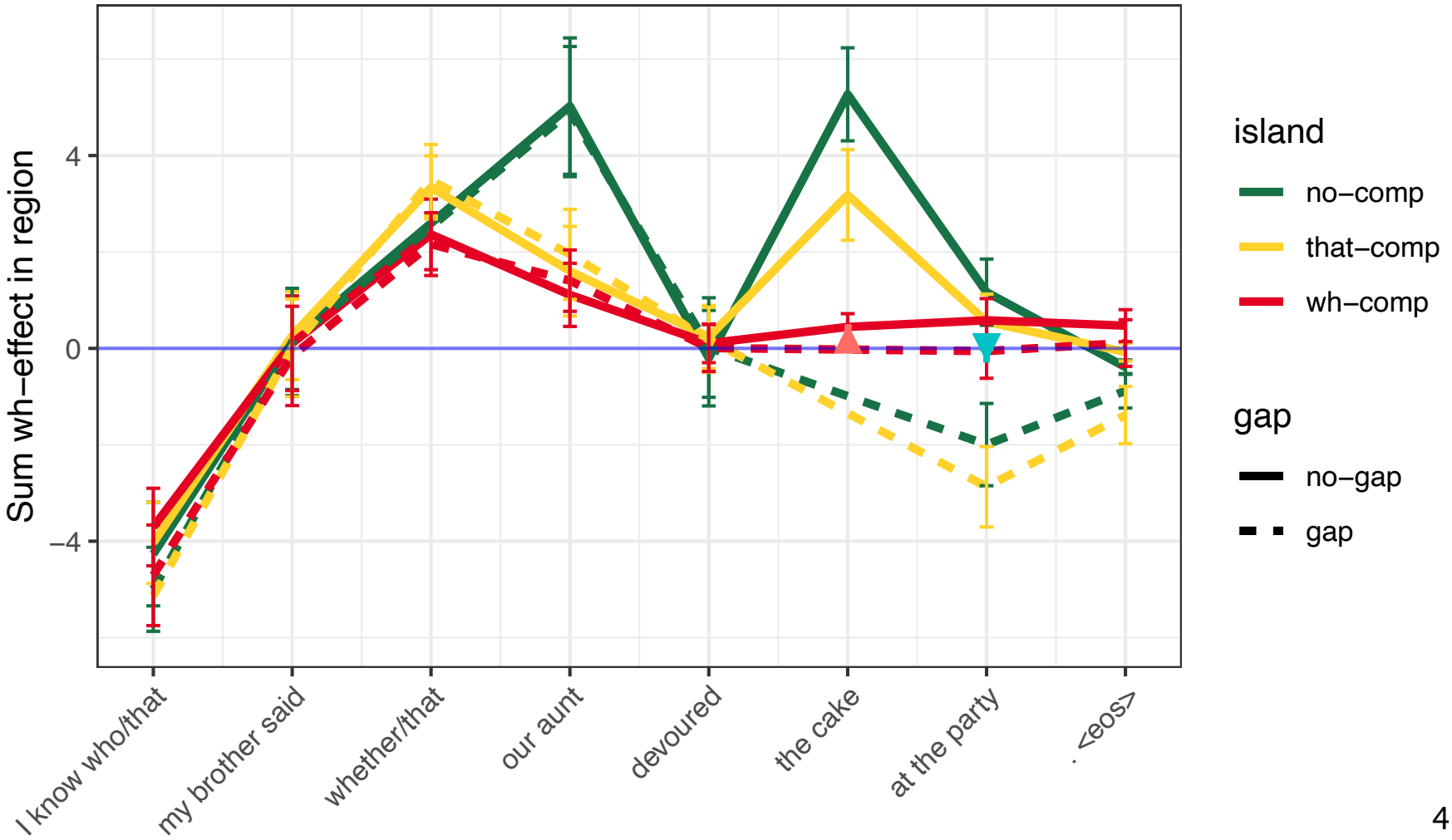


✓ *I know that my brother said **whether** our aunt devoured the cake at the party.*

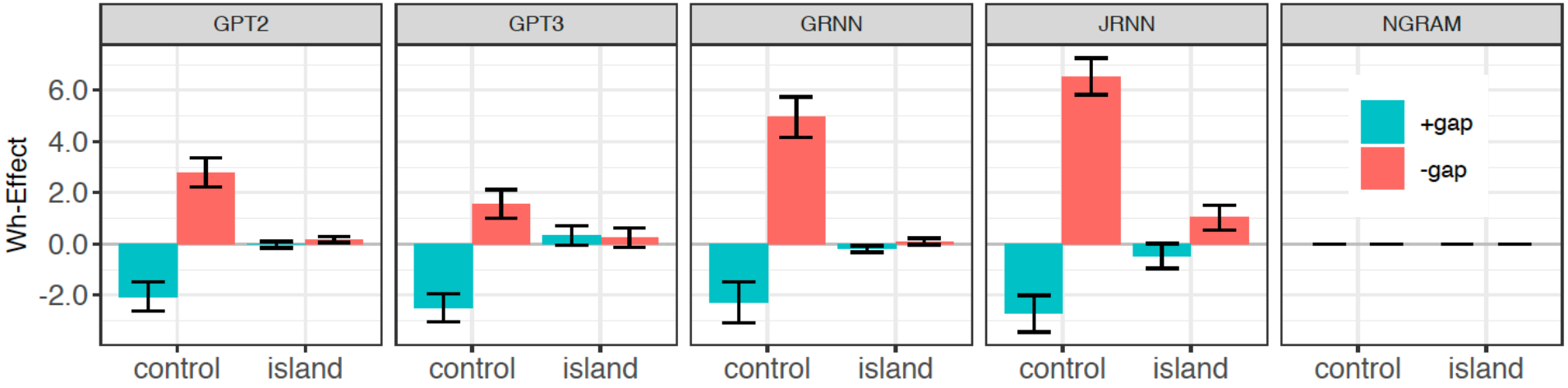
* *I know what my brother said **whether** our aunt devoured the cake at the party.*

* *I know that my brother said **whether** our aunt devoured _____ at the party.*

* *I know what my brother said **whether** our aunt devoured _____ at the party.*



Results for **WH-islands**



(Wilcox et al., in prep)

Potential concern #2

Could deep-learning models have difficulty threading ***any*** type of expectation into a syntactic island?

Gendered-pronoun Expectation Control

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

Gender Expectation Effect (#-✓ should be *positive)*

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

Gender Expectation Effect (#-✓ should be *positive*)

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

Gender Expectation Effect (#-✓ should be *positive*)

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

\# *The actress said whether they insulted* **his** *friends.*
[ISLAND, MISMATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

Gender Expectation Effect (#-✓ should be *positive*)

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

?

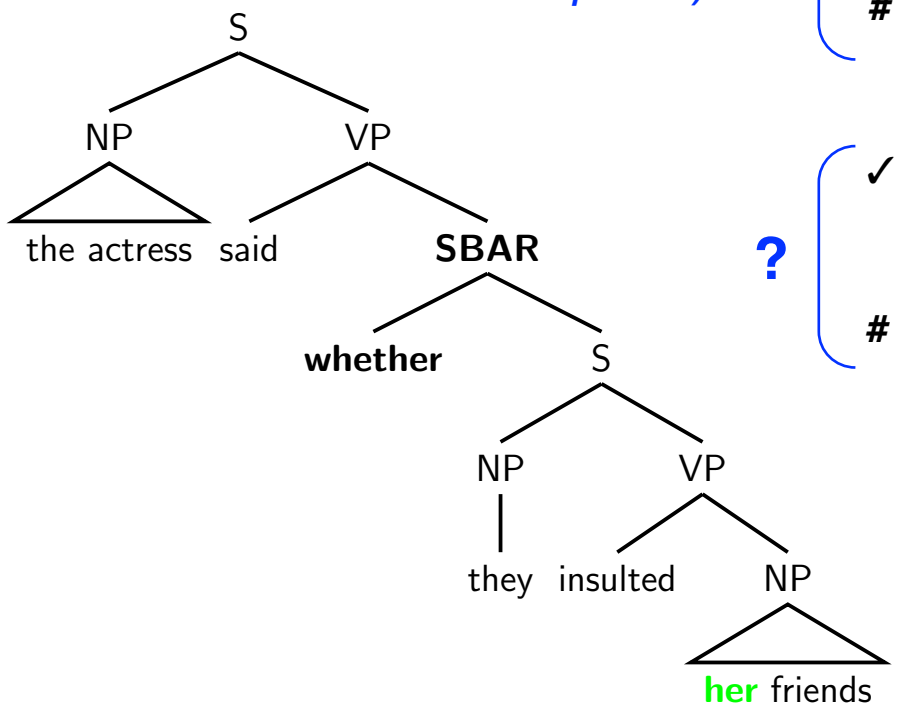
✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

\# *The actress said whether they insulted* **his** *friends.*
[ISLAND, MISMATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.



Gender Expectation Effect (#-✓ should be *positive*)

?

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

\# *The actress said whether they insulted* **his** *friends.*
[ISLAND, MISMATCH]

(Wilcox et al., 2019; in prep)

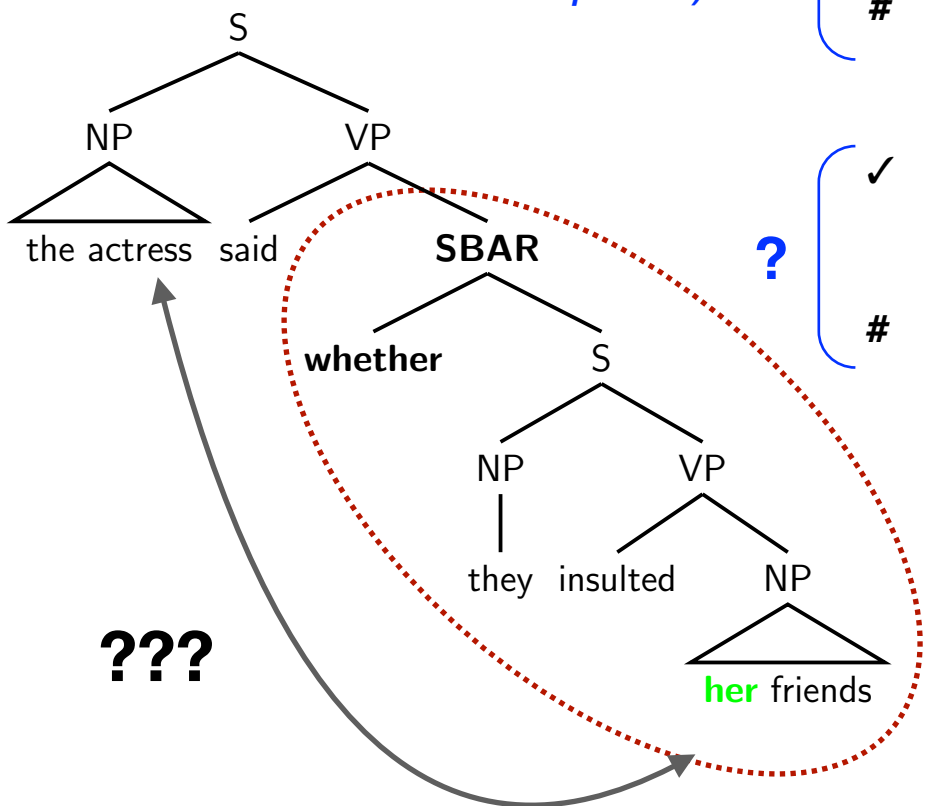
Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

Gender Expectation Effect (#-✓ should be *positive*)

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]



✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

\# *The actress said whether they insulted* **his** *friends.*
[ISLAND, MISMATCH]

(Wilcox et al., 2019; in prep)

Gendered-pronoun Expectation Control

- Worry: Can the models thread **any** expectation into islands?
- Test with expectation for **gendered pronouns** set up by **culturally or morphologically gendered subjects**.

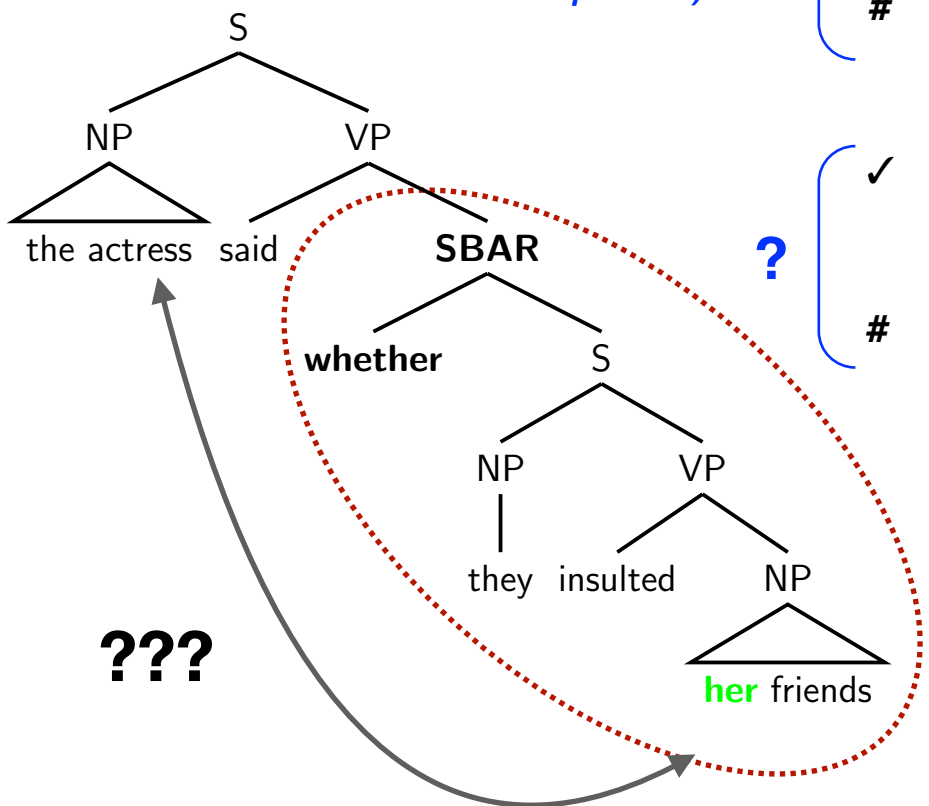
Gender Expectation Effect (#-✓ should be *positive*)

✓ *The actress said that they insulted* **her** *friends.*
[CONTROL, MATCH]

\# *The actress said that they insulted* **his** *friends.*
[CONTROL, MISMATCH]

✓ *The actress said whether they insulted* **her** *friends.*
[ISLAND, MATCH]

\# *The actress said whether they insulted* **his** *friends.*
[ISLAND, MISMATCH]



If models can thread gender expectation into islands, the gender expectation effect should **look the same in islands as in the control conditions.**

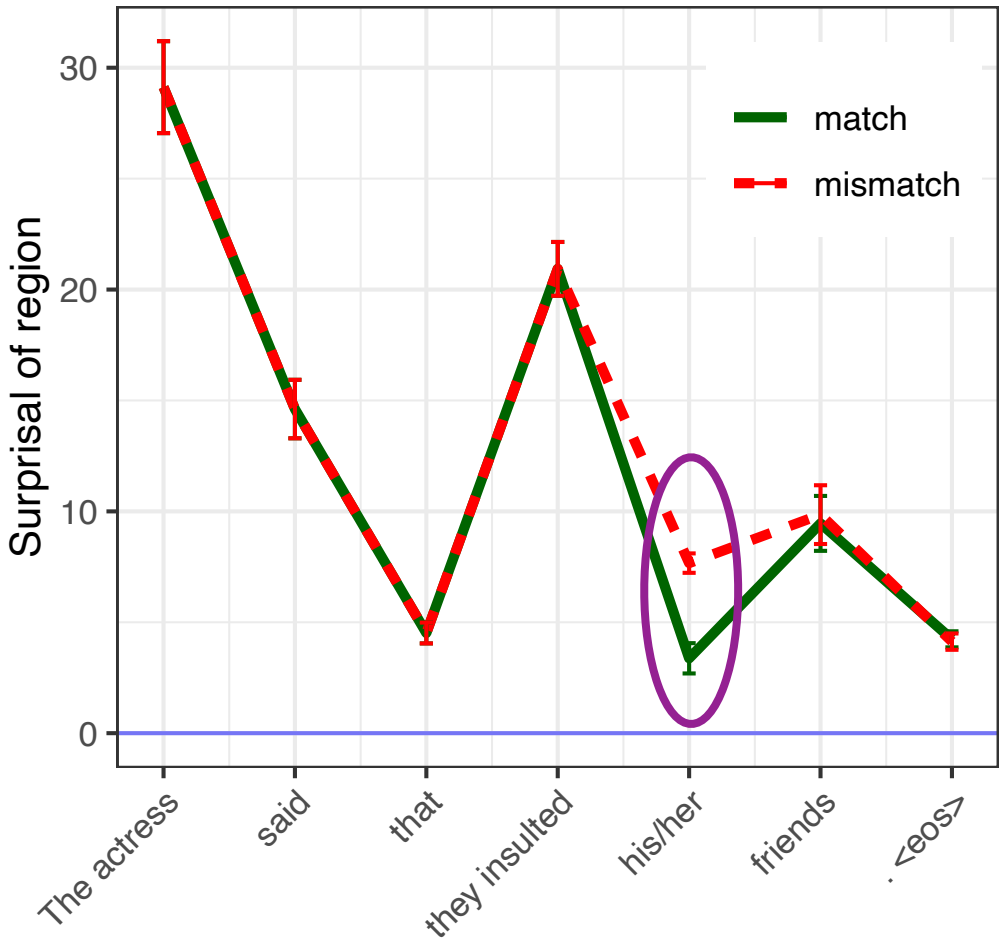
(Wilcox et al., 2019; in prep)

*The actress said **that** they insulted her friends.*

*The actress said **that** they insulted his friends.*

*The actress said **that** they insulted her friends.*

*The actress said **that** they insulted his friends.*

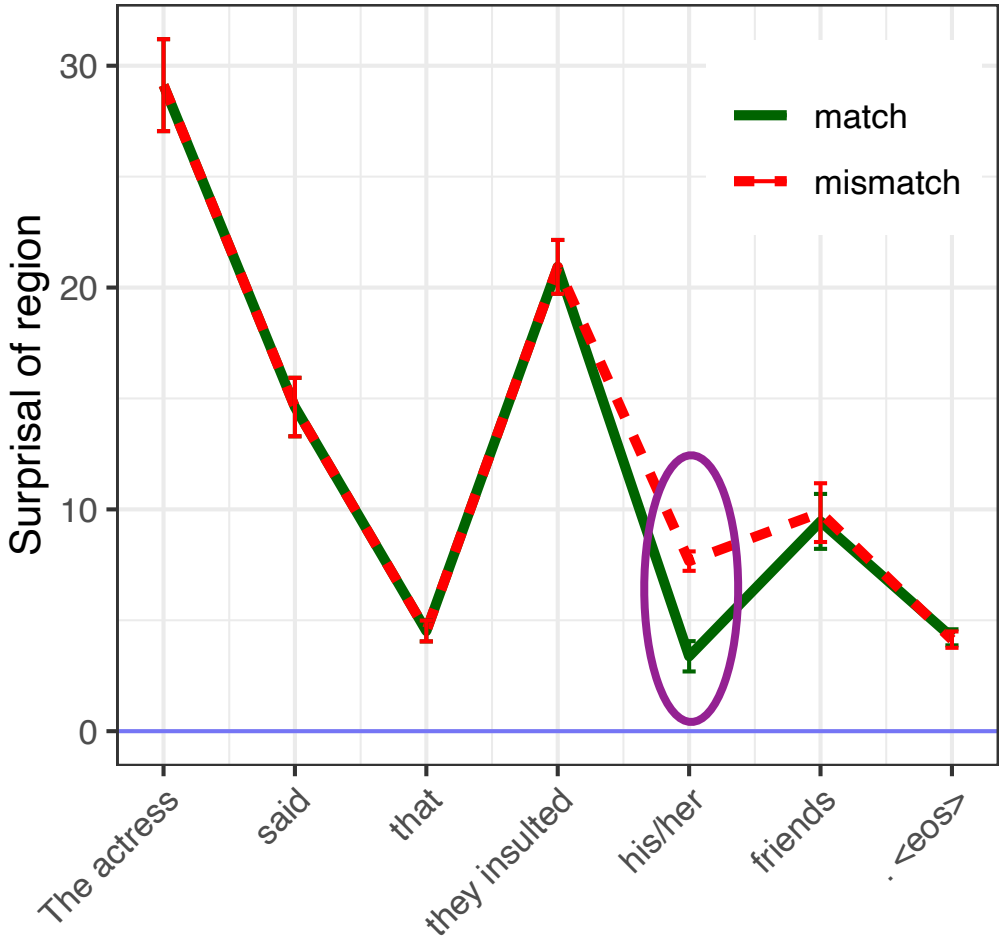


*The actress said **that** they insulted her friends.*

*The actress said **that** they insulted his friends.*

*The actress said **whether** they insulted her friends.*

*The actress said **whether** they insulted his friends.*

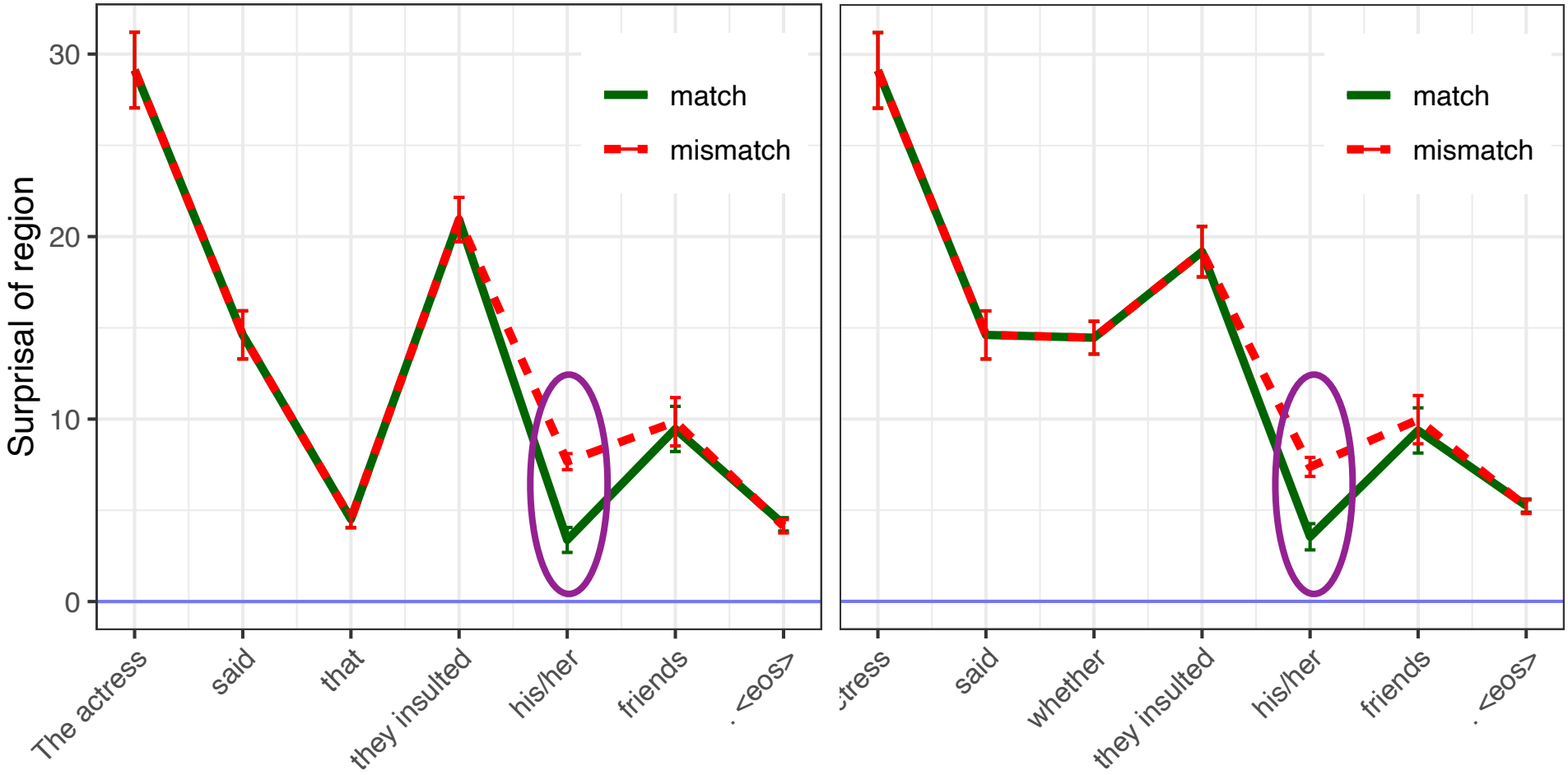


*The actress said **that** they insulted her friends.*

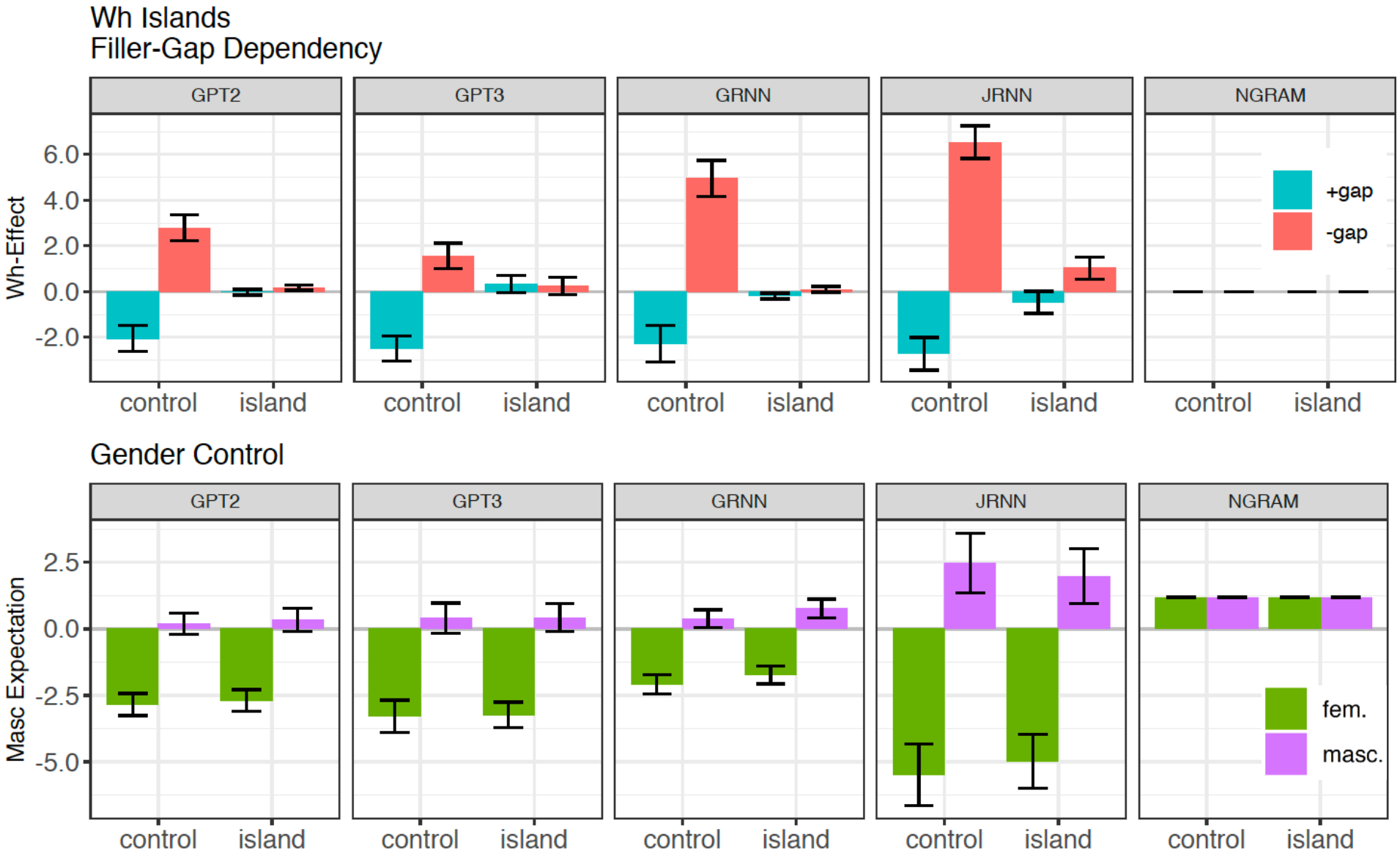
*The actress said **that** they insulted his friends.*

*The actress said **whether** they insulted her friends.*

*The actress said **whether** they insulted his friends.*



Filler–gap vs. gender expectations in WH-islands



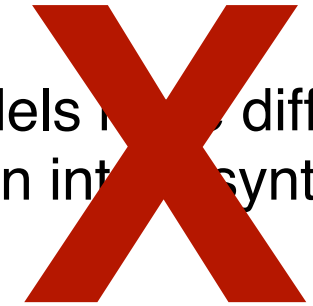
(Wilcox et al., in prep)

Potential concern #2

Could deep-learning models have difficulty threading ***any*** type of expectation into a syntactic island?

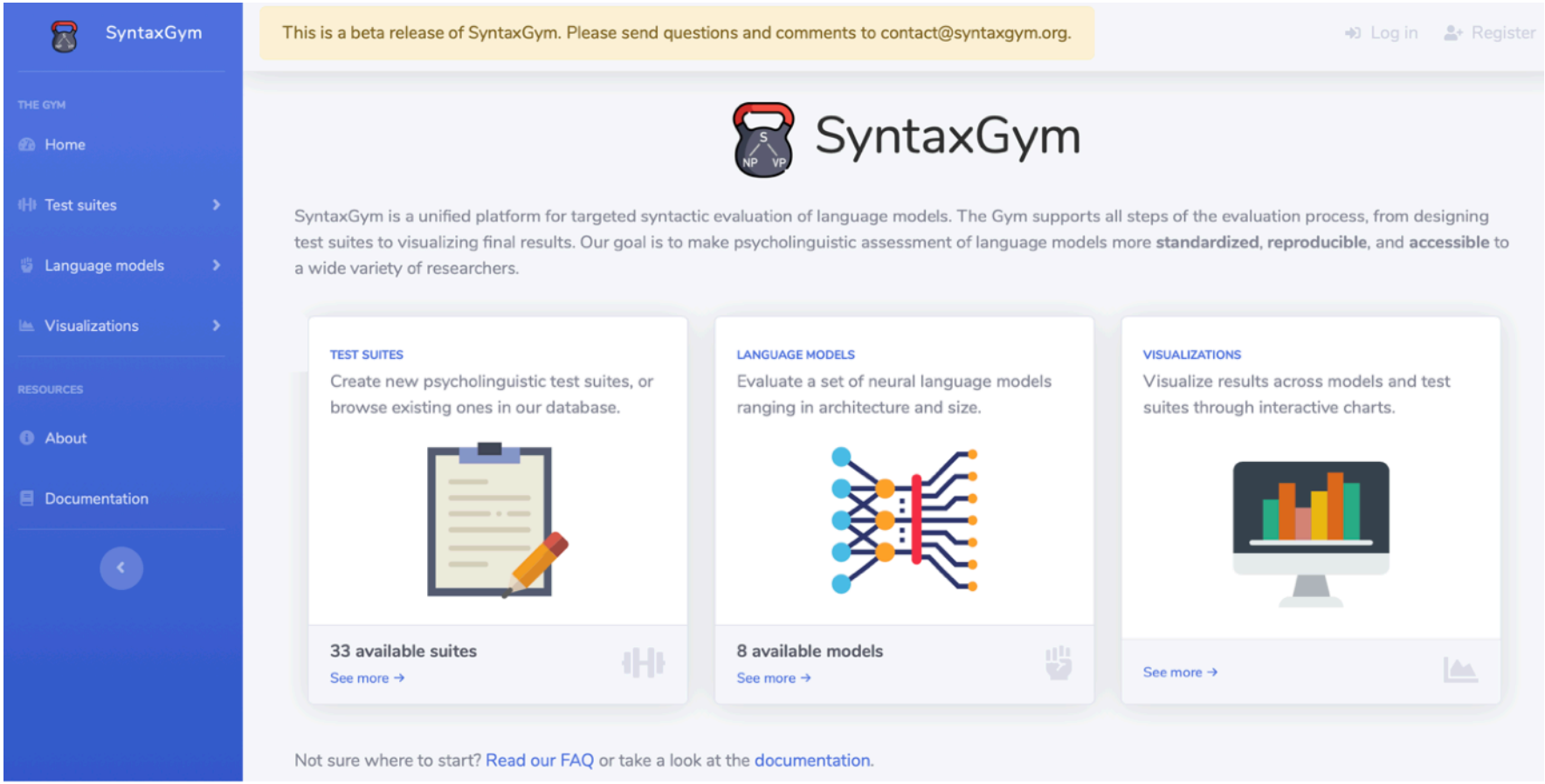
Potential concern #2 — *addressed*

Could deep-learning models have difficulty threading ***any*** type of expectation into a syntactic island?



Deep-learning models that learn island constraints still propagate pronoun gender expectations into islands

Psycholinguistic tests of AI language models

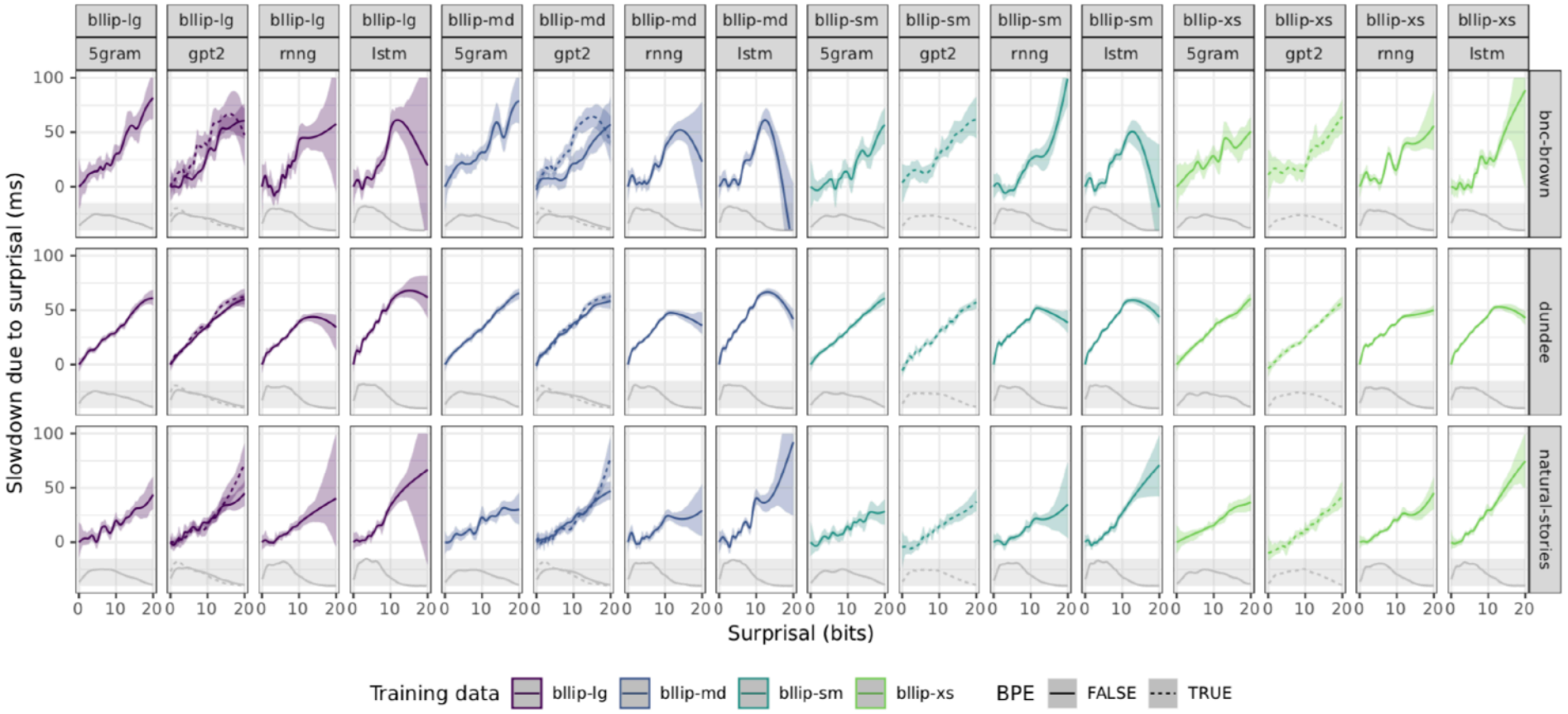


http://syntaxgym.org

(Gauthier et al., 2020)

Quantitative calibration to human processing

- The surprisal–RT relationship in naturalistic reading:



(Wilcox et al., 2020)

Low-tech, crowd-sourceable reading

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

F J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

The x-x-x

F J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

of

F

dog

J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

of dog

F J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

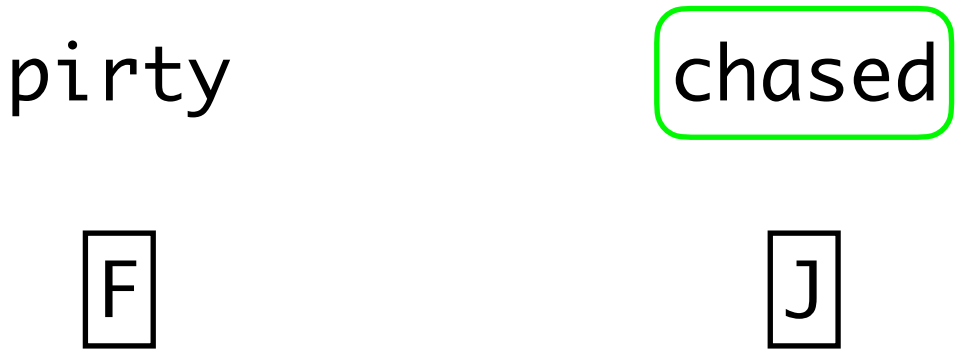
- The **maze task**
- Choose the string that fits given the preceding context

pirty	chased
F	J

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context



(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

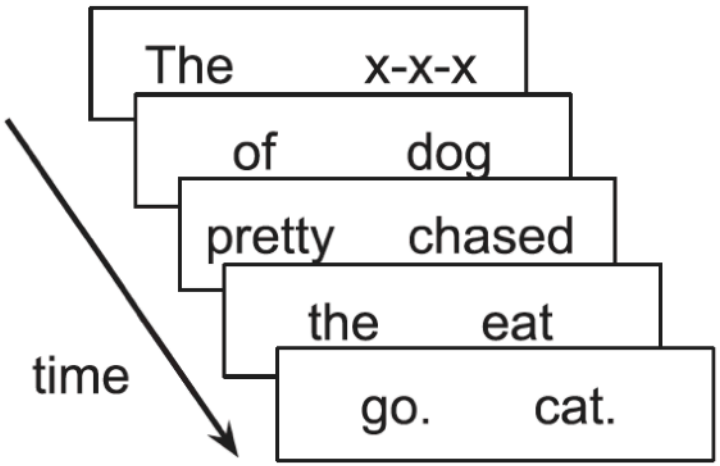
(Forster et al., 2009; Boyce et al., 2020)

Low-tech, crowd-sourceable reading

- The **maze task**
- Choose the string that fits given the preceding context

the eat

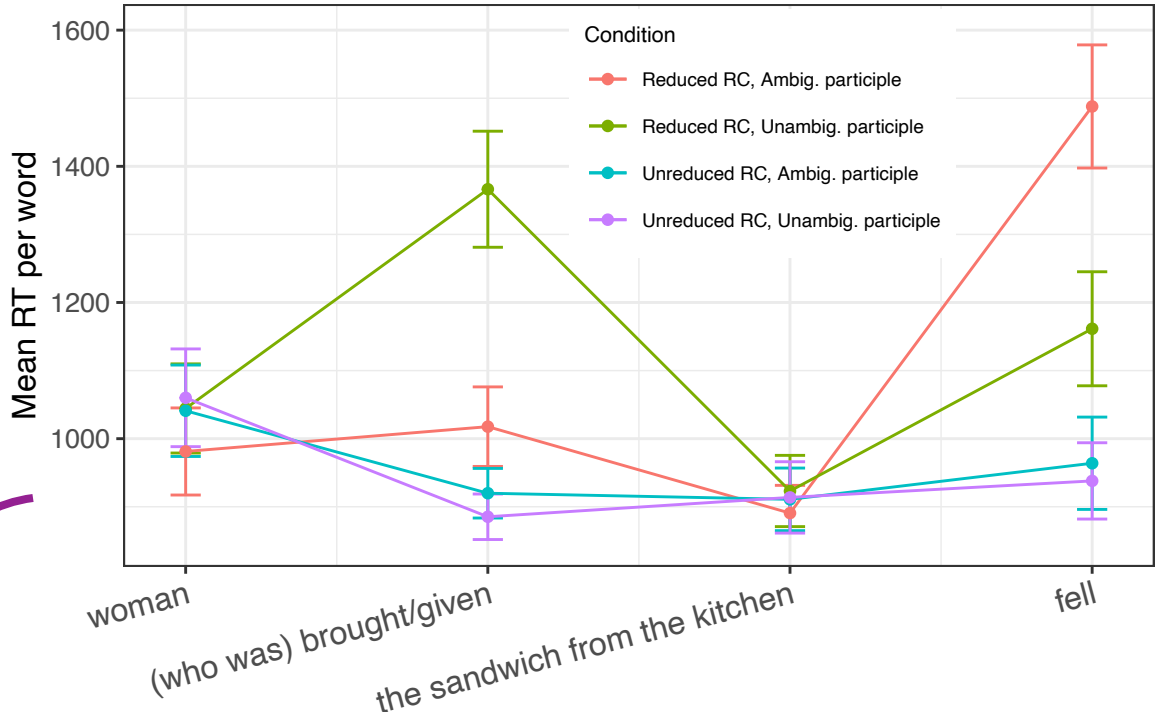
F J



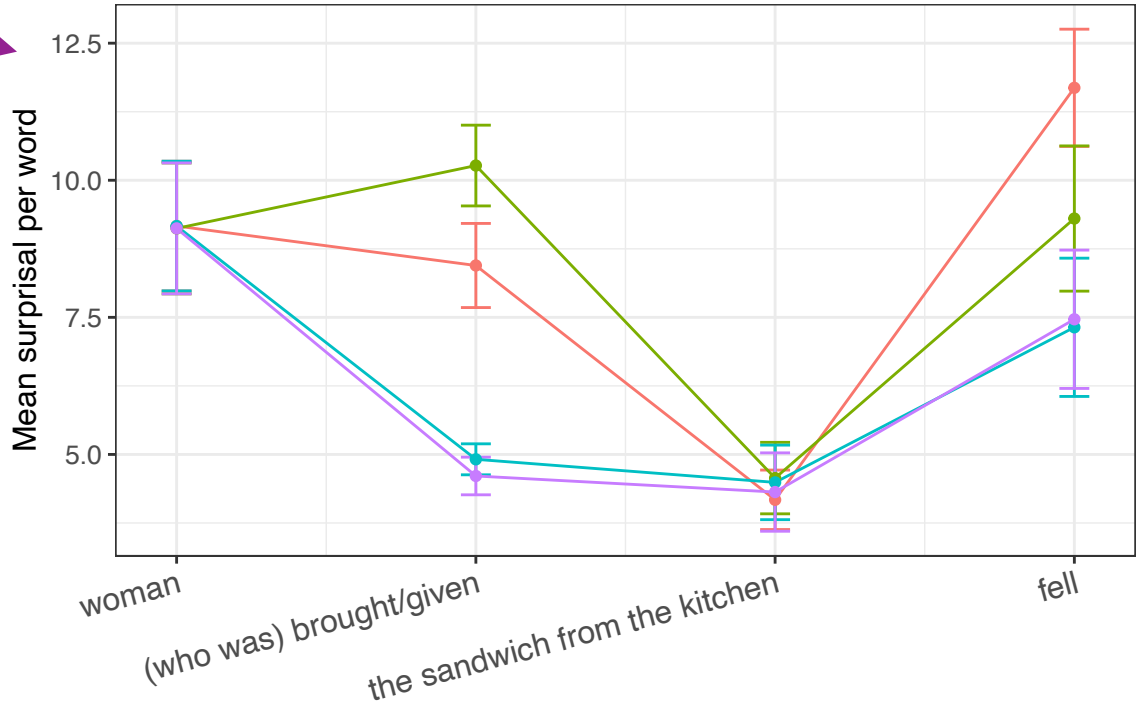
(Forster et al., 2009; Boyce et al., 2020)

Human reaction times

*Pooling many controlled experiments, regress human RTs against model surprisal and examine **residual***



GPT-2 Surprisal



(Wilcox et al., 2021; see also van Schijndel & Linzen, 2020)

✓ *I know that ... the CEO showed the slides to the guests after lunch.*

✳ *I know what ... the CEO showed the slides to the guests after lunch.*

✳ *I know that ... the CEO showed _____ to the guests after lunch.*

✓ *I know what ... the CEO showed _____ to the guests after lunch.*

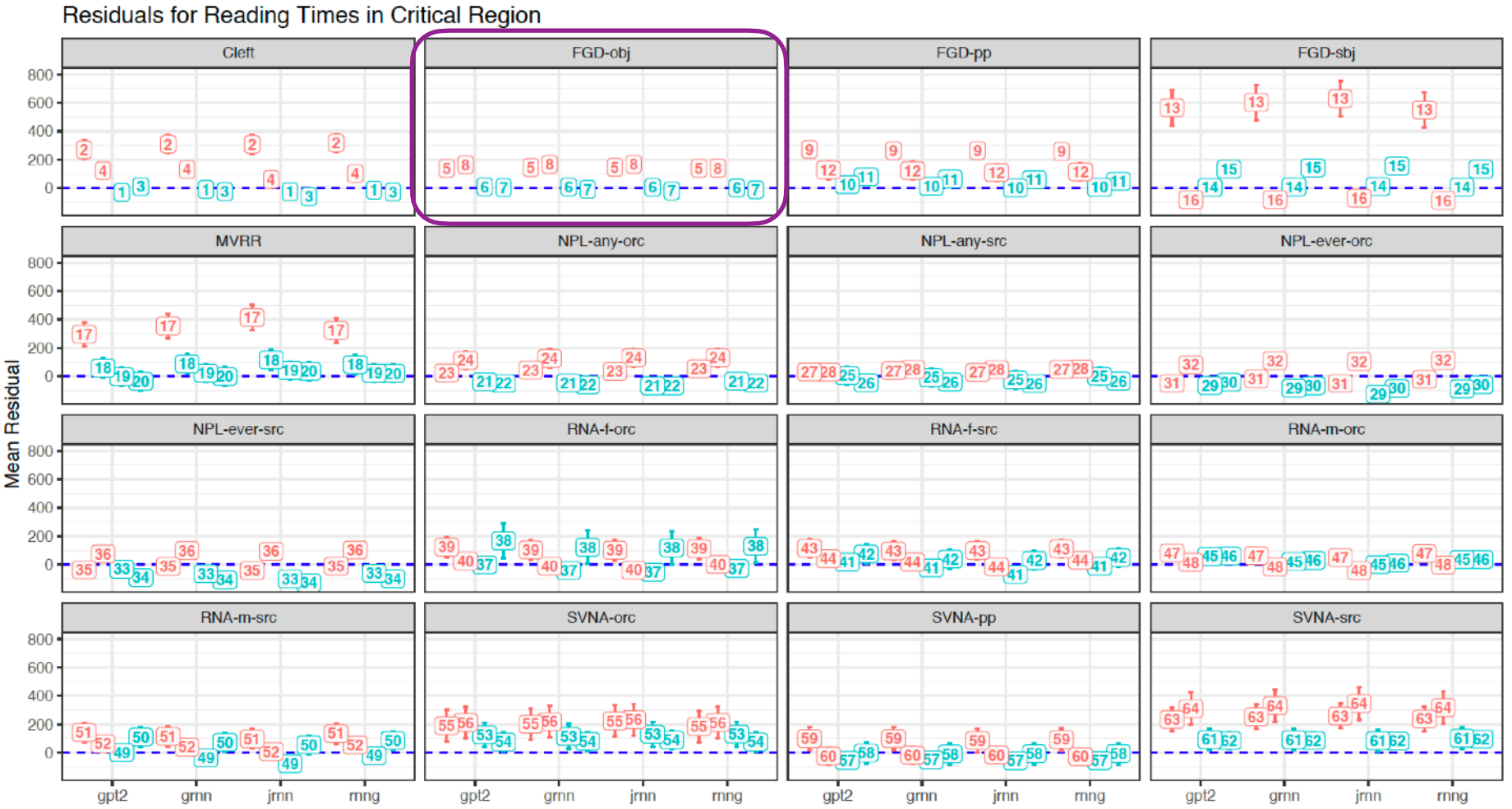
(Wilcox et al., 2021)

✓ *I know that ... the CEO showed the slides to the guests after lunch.*

∗ *I know what ... the CEO showed the slides to the guests after lunch.*

∗ *I know that ... the CEO showed _____ to the guests after lunch.*

✓ *I know what ... the CEO showed _____ to the guests after lunch.*



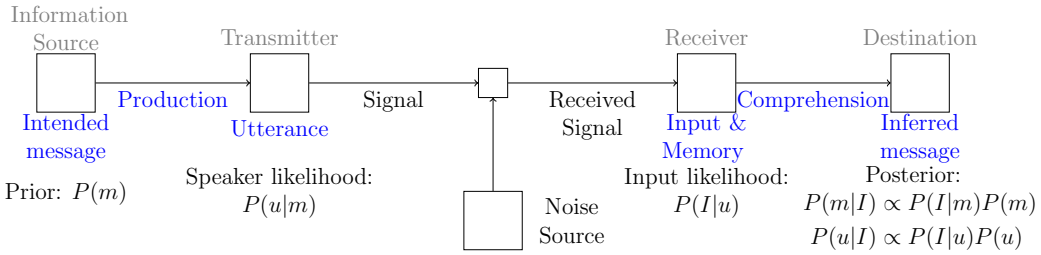
(Wilcox et al., 2021)

Summary: what I have argued today

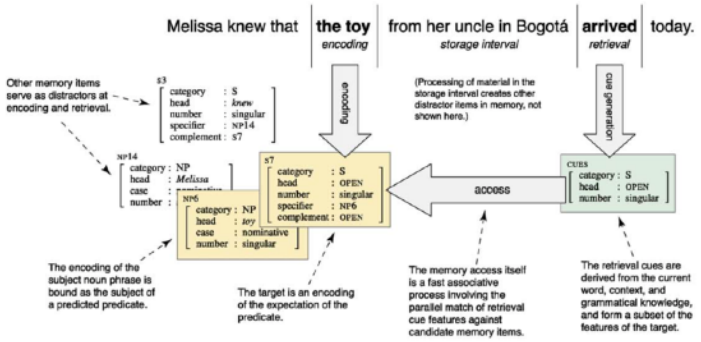
- Standard deep-learning models **learn remarkably subtle features of human grammar** from a **childhood's worth of linguistic input** (no real-world grounding needed!)
- However, these models' predictions are **not quantitatively aligned with human comprehension behavior when expectations about grammatical structure are violated**
- Deep-learning models offer **insights into learnability** and **a powerful scientific tool for expectation estimation,** but **not a theoretical account** of human language representation and processing

Other ingredients for theory of human language comprehension

- Noisy-channel mechanisms for error detection & robustness (Levy 2008, Gibson et al., 2013, Futrell et al., 2020)

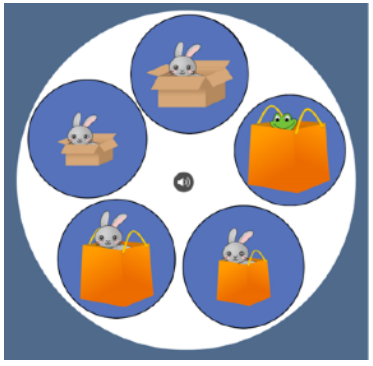


- Limitations on fidelity of memory representations & access (Lewis et al., 2006)



- Incremental semantic representations evaluable in context (Jacobson 1999, Aparicio et al. in prep)

Click on the rabbit in the big...



Mary loves and John hates...

λx[LOVE(x)(mary) ∧HATE(x)(john)]

Acknowledgments

- **Collaborators**: Miguel Ballesteros, Veronica Boyce, Richard Futrell, Jon Gauthier, Jennifer Hu, Takashi Morita, Peng Qian, Pranali Vani, Ethan Wilcox
- **Funders:** National Science Foundation, MIT–IBM AI Research Lab, MIT Quest for Intelligence, Google, Elemental Cognition
- The **MIT Computational Psycholinguistics Laboratory**

Thank you for listening!

http://syntaxgym.org

http://cpl.mit.edu

http://www.mit.edu/~rplevy

References

Boyce, V., Futrell, R., & Levy, R. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. Journal of Memory and Language, 111, 1–13.
Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Proceedings of NeurIPS.
Dyer, C., Kuncoro, A., Ballesteros, M., & Smith, N. A. (2016). Recurrent neural network grammars. Proceedings of NAACL.
Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 20, 641–655.
Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. Behavior Research Methods, 41(1), 163–171.
Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state, In Proceedings of NAACL.
Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. Cognitive science, 44(3), e12814.
Gauthier, J., Hu, J., Wilcox, E., Qian, P., & Levy, R. P. (2020). SyntaxGym: An online platform for targeted evaluation of language models, In Proceedings of the 58th annual meeting of the Association for Computational Linguistics.
Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. Proceedings of the National Academy of Sciences, 110(20), 8051-8056.
Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. Proceedings of NAACL 2018.
Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model, In Proceedings of NAACL.
Hu, J., Gauthier, J., Qian, P., Wilcox, E., & Levy, R. P. (2020). A systematic assessment of syntactic generalization in neural language models, In Proceedings of the 58th annual meeting of the Association for Computational Linguistics.
Jacobson, P. (1999). Towards a variable-free semantics. Linguistics and philosophy, 117-184.
Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. In International conference on machine learning (pp. 2342-2350). PMLR.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203–205.
Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126–1177.
Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In Proceedings of the 2008 conference on empirical methods in natural language processing (pp. 234-243).
Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. Trends in cognitive sciences, 10(10), 447-454.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. Psychonomic Bulletin & Review, 3(4), 504-509.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3), 302–319.
Van Schijndel, M., & Linzen, T. (2020). Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. PsyArXiv.
Wilcox, E., Levy, R. P., & Futrell, R. (2019). What syntactic structures block dependencies in RNN language models? In Proceedings of the 41st annual meeting of the Cognitive Science Society.
Wilcox, E., Levy, R. P., Morita, T., & Futrell, R. (2018). What do RNN language models learn about filler--gap dependencies? In Proceedings of the workshop on analyzing and interpreting neural networks for NLP.
Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. P. (2020). On the predictive power of neural language models for human real-time comprehension behavior, In Proceedings of the 42nd annual meeting of the Cognitive Science Society.
Wilcox, E. G., Vani, P., & Levy, R. P. (2021). A Targeted Assessment of Incremental Processing in Neural LanguageModels and Humans. Proceedings of ACL 2021.