



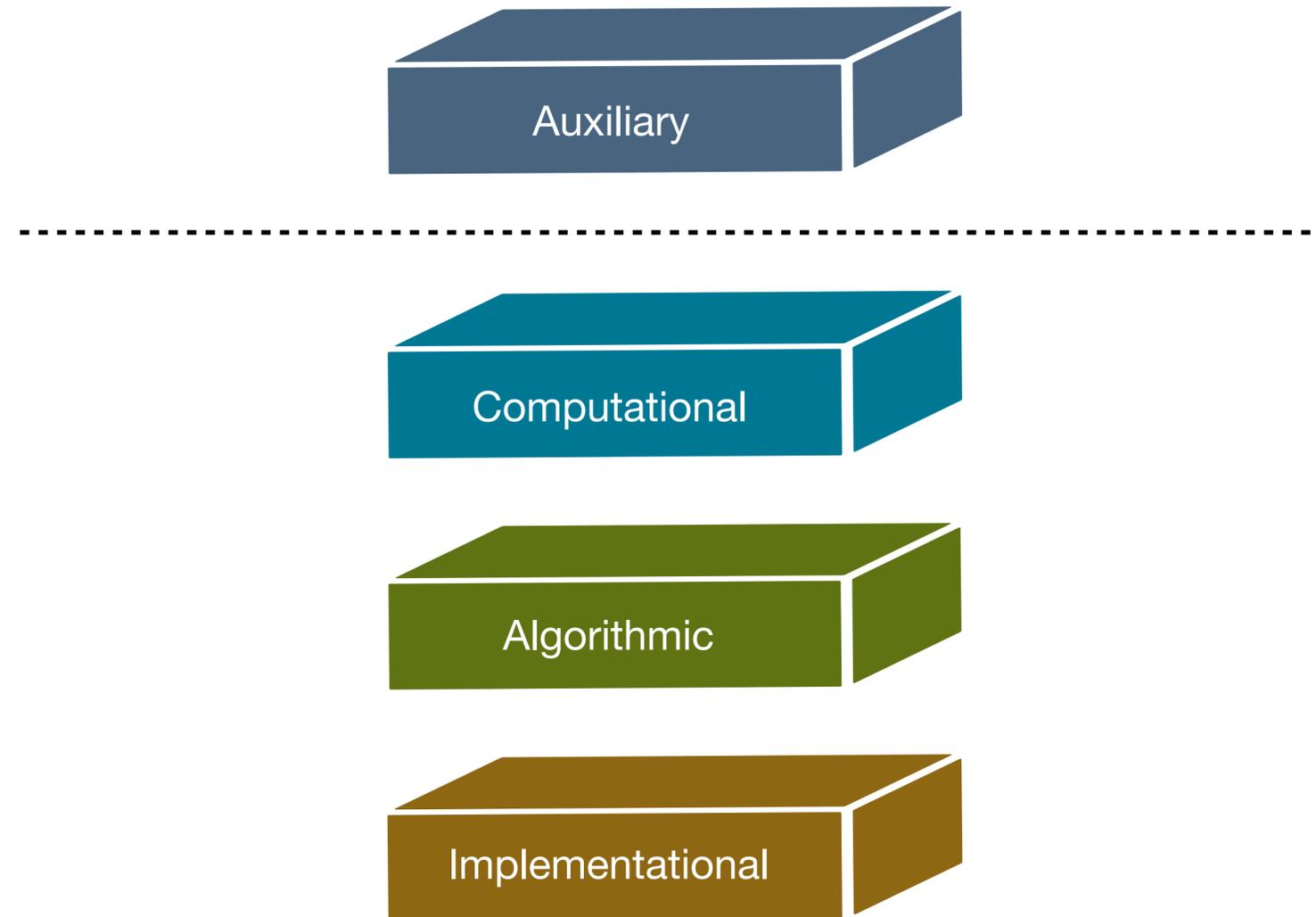
PLIN0033
27 March 2026

When Psycholinguistics meets Computational Linguistics

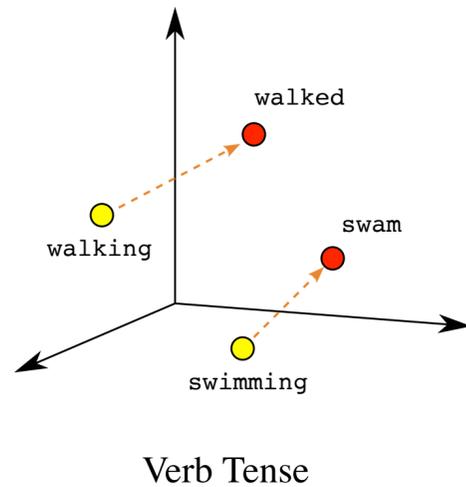
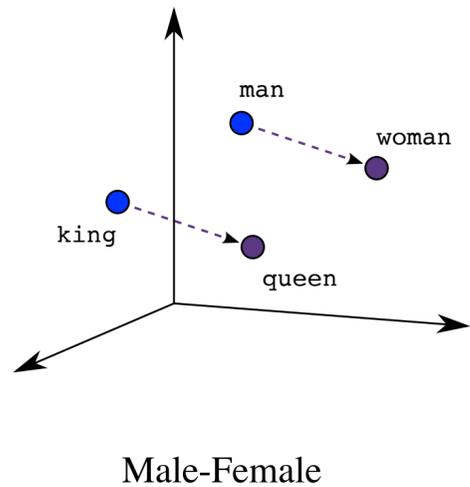
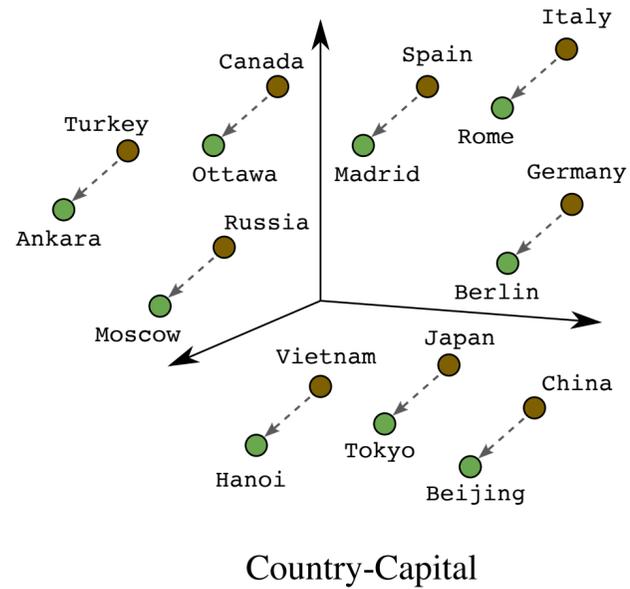
Surprisal Theory and Beyond

Mario Giulianelli

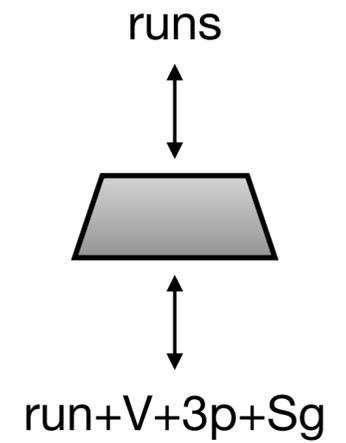
What is the role for computational models in linguistic theory?



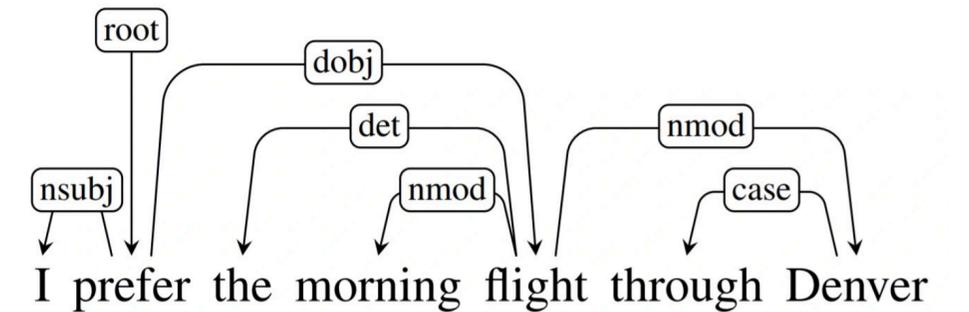
What is the role for computational models in linguistic theory?



Distributional word
and sentence representations

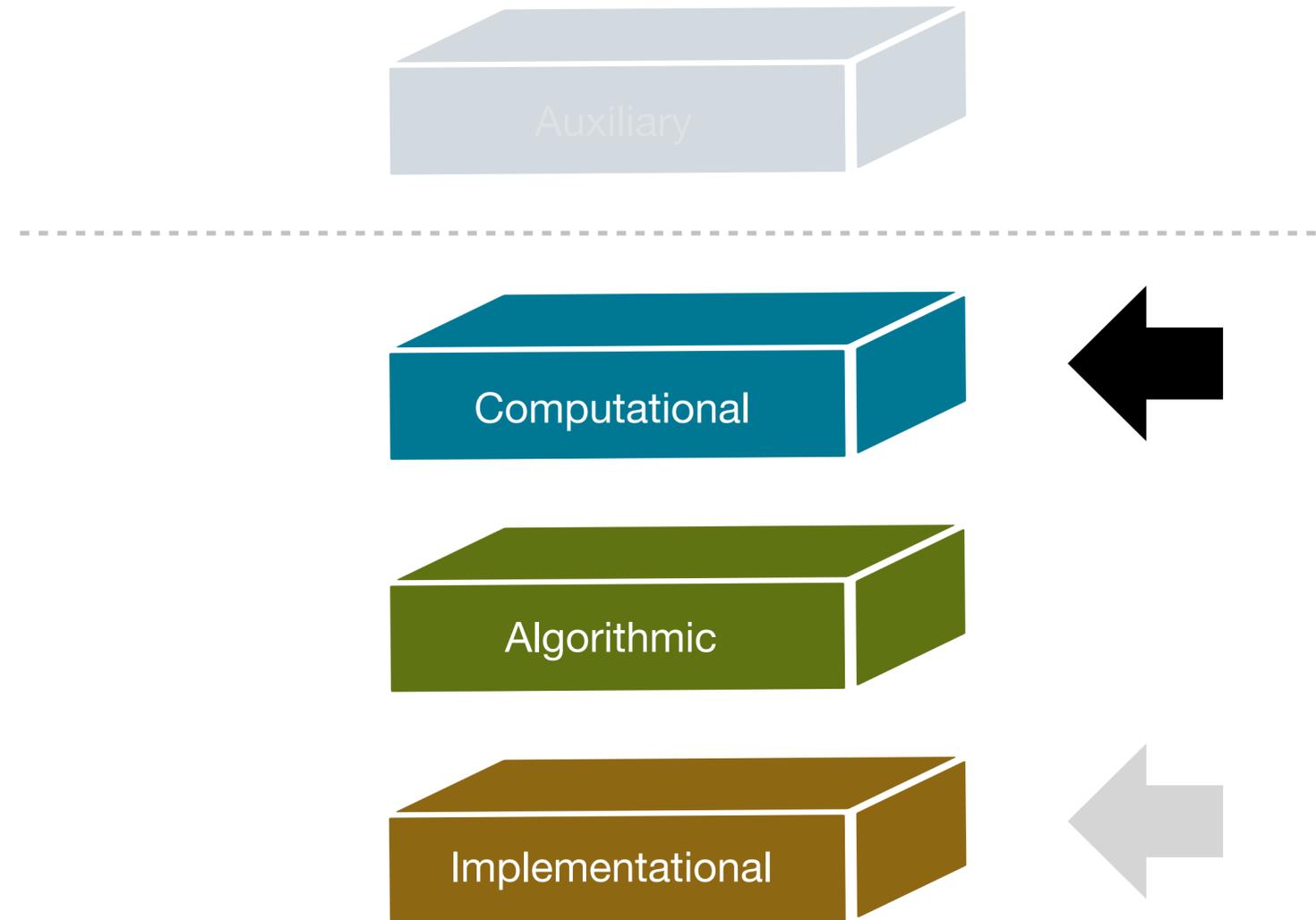


Morphological analyser

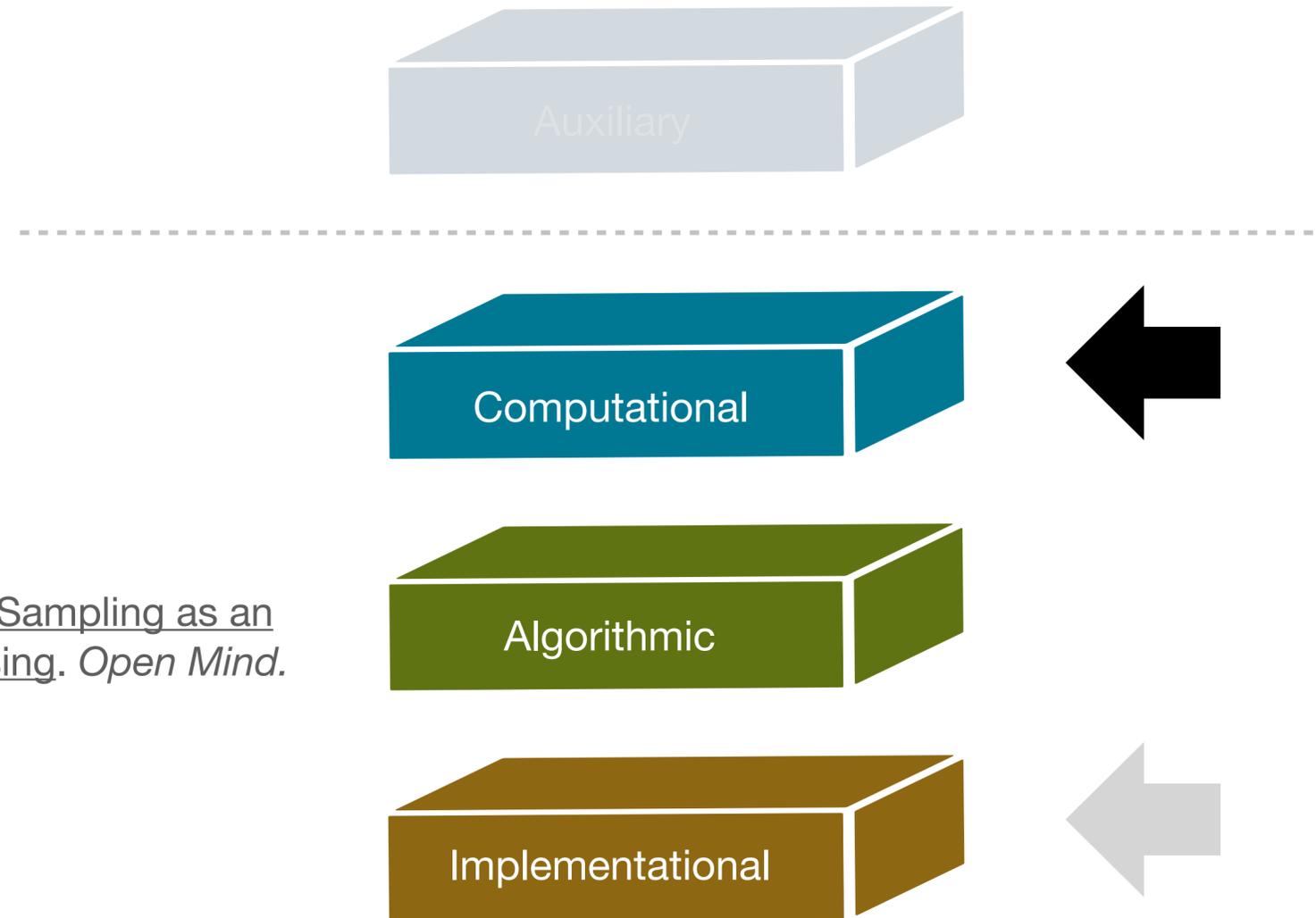


Parsers

What is the role for computational models in linguistic theory?



What is the role for computational models in linguistic theory?



Hoover et al. (2023). The Plausibility of Sampling as an Algorithmic Theory of Sentence Processing. *Open Mind*.

What is the role for computational models in linguistic theory?

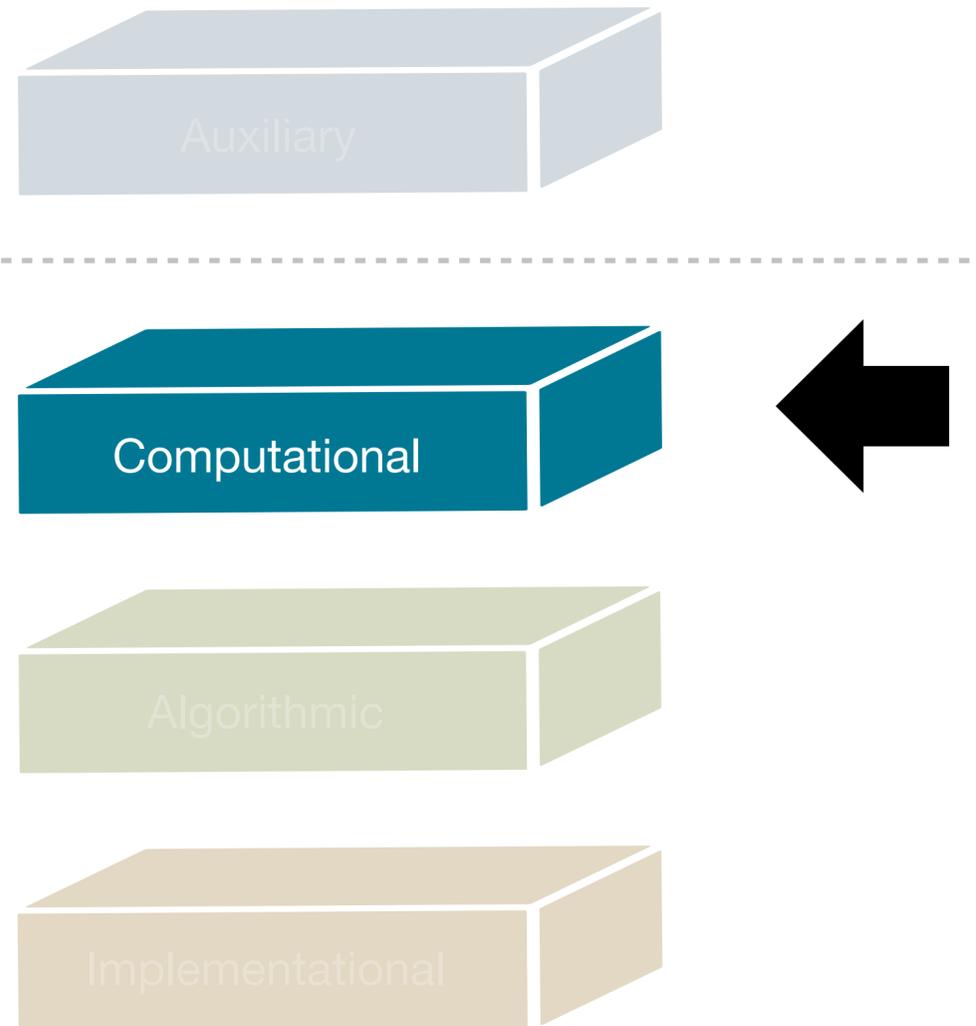
“The most abstract is the level of **what** the device does and **why.**”

“[a] level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is **independent of the particular mechanisms and structures that implement them in our heads.**

[...]

the analysis of the problem as an **information processing task.**”

“mapping from one kind of information to another”



The *What* and the *Why*

Non-incremental Stimulus

$$w_{1:T}$$

BBC wants to change its viewers' metabolism

Incremental Stimulus

$$w_{<t}$$

$$w_t$$

BBC wants to change its viewers' metabolism

Stimulus

The *What* and the *Why*

Non-incremental Stimulus



BBC wants to change its viewers' metabolism

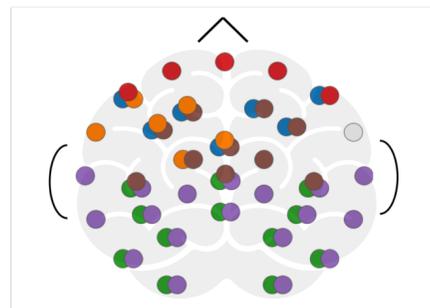
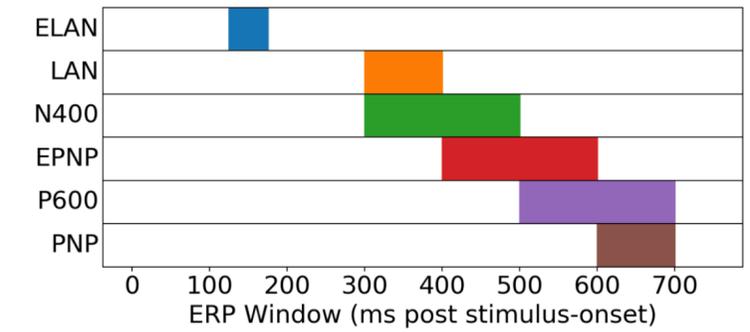
Incremental Stimulus



BBC wants to change its viewers' metabolism

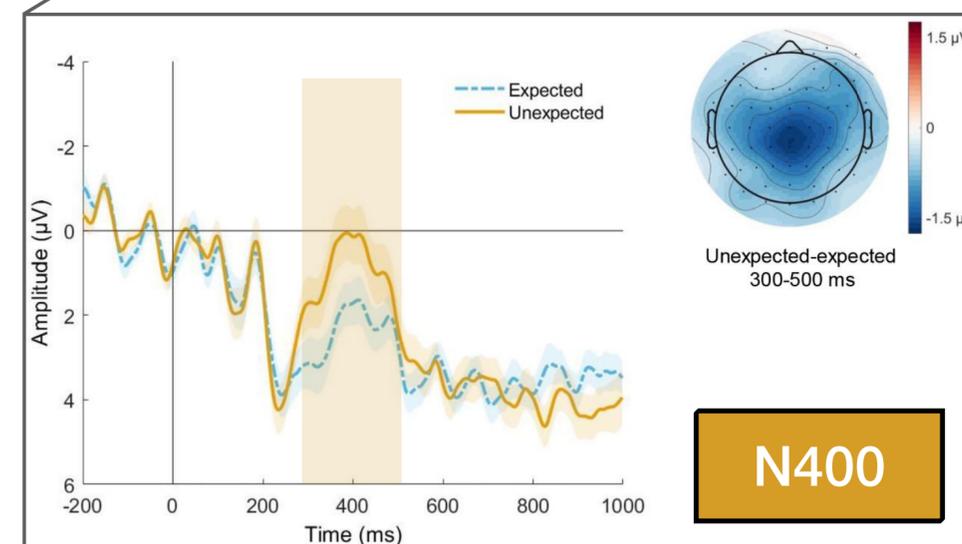
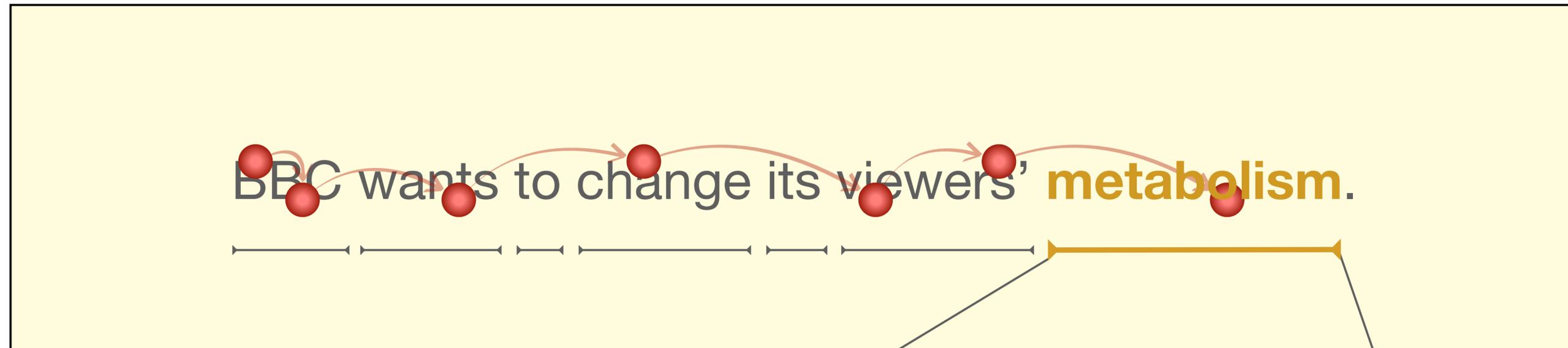
Stimulus

E.g., EEG response



Response

Behavioural and neural responses provide a window into the information processing tasks carried out during language processing.



The *What* and the *Why*



The *What* and the *Why*

Hypothesis
(Explanation)

Stimulus

$w_{<t}$

w_t

$t \in 1..T$

Response

e.g.

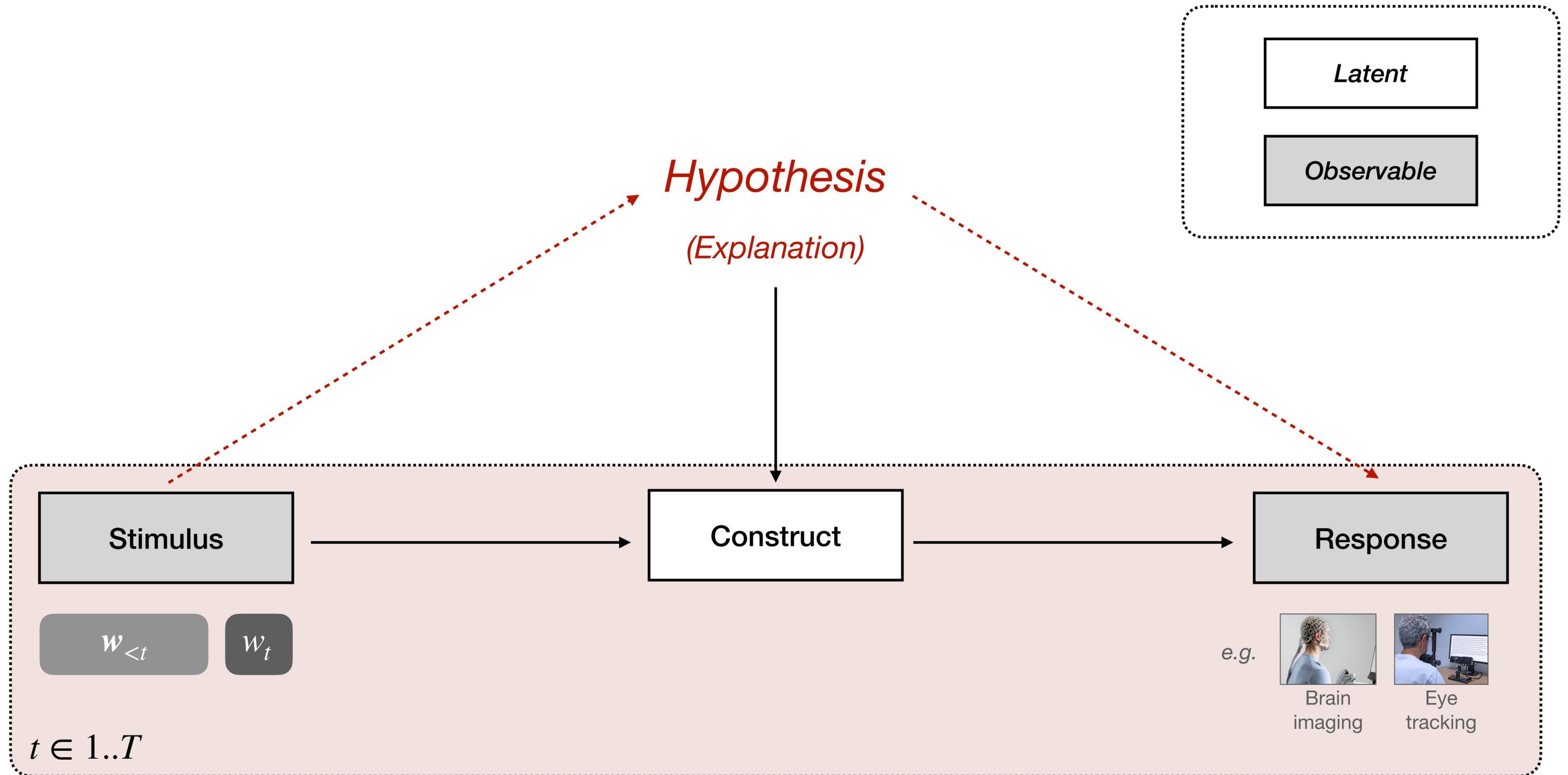


Brain
imaging

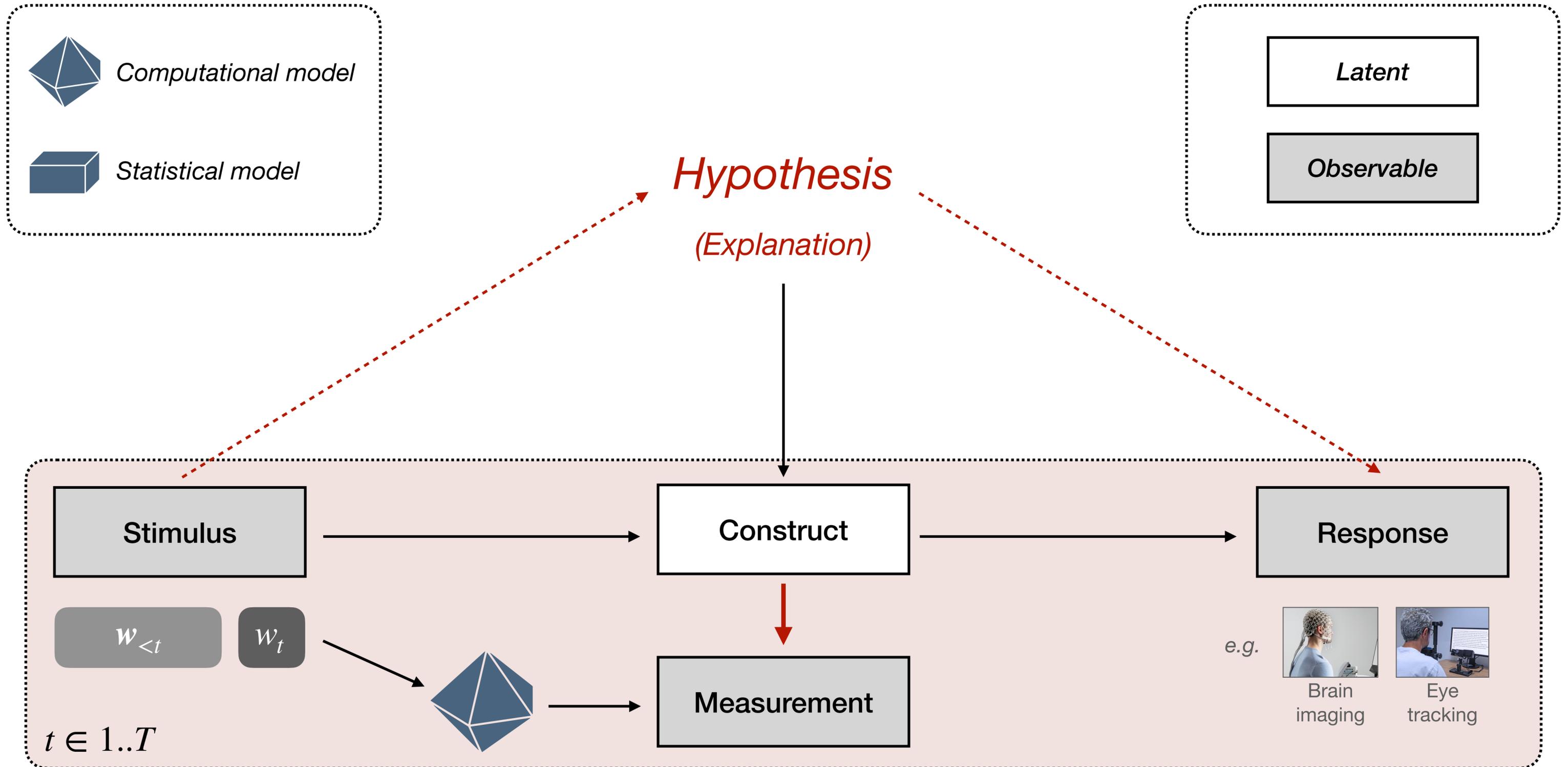


Eye
tracking

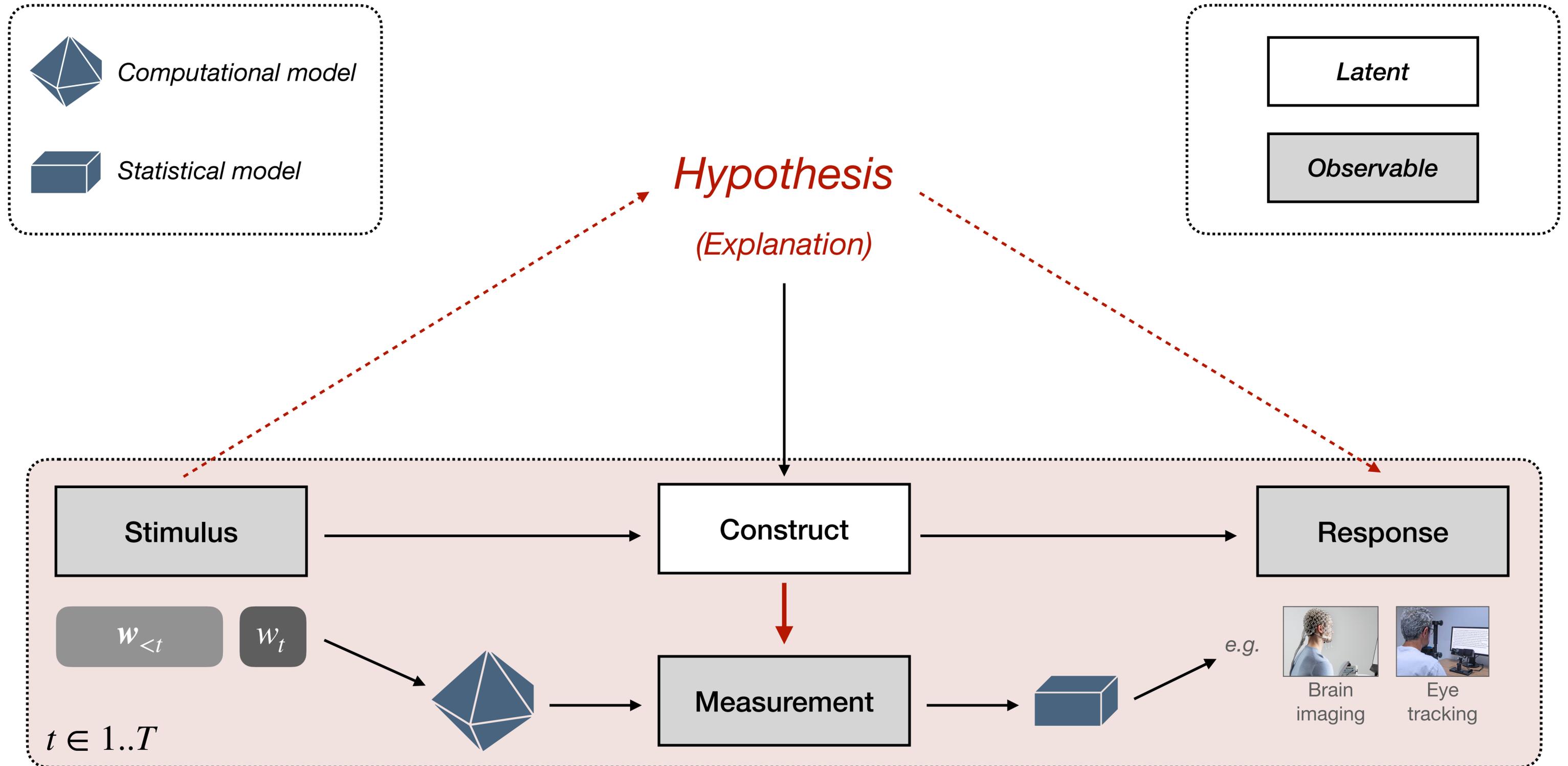
The *What* and the *Why*



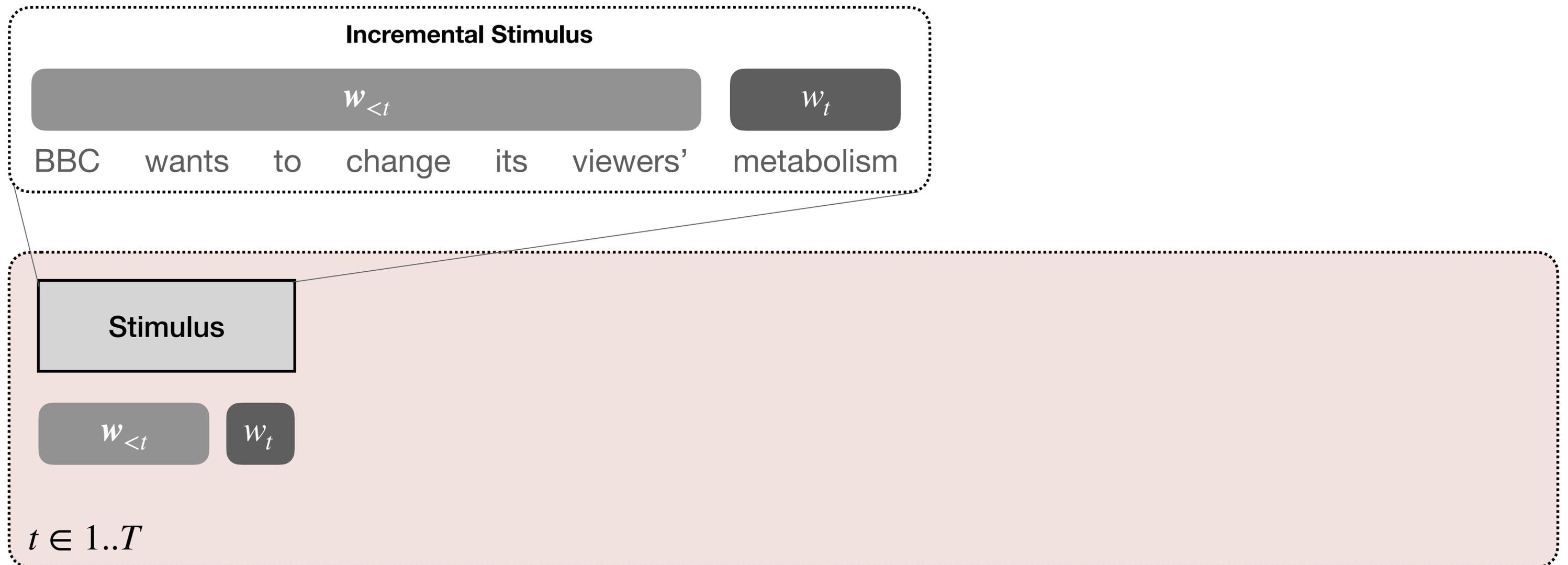
The *What* and the *Why*



The *What* and the *Why*

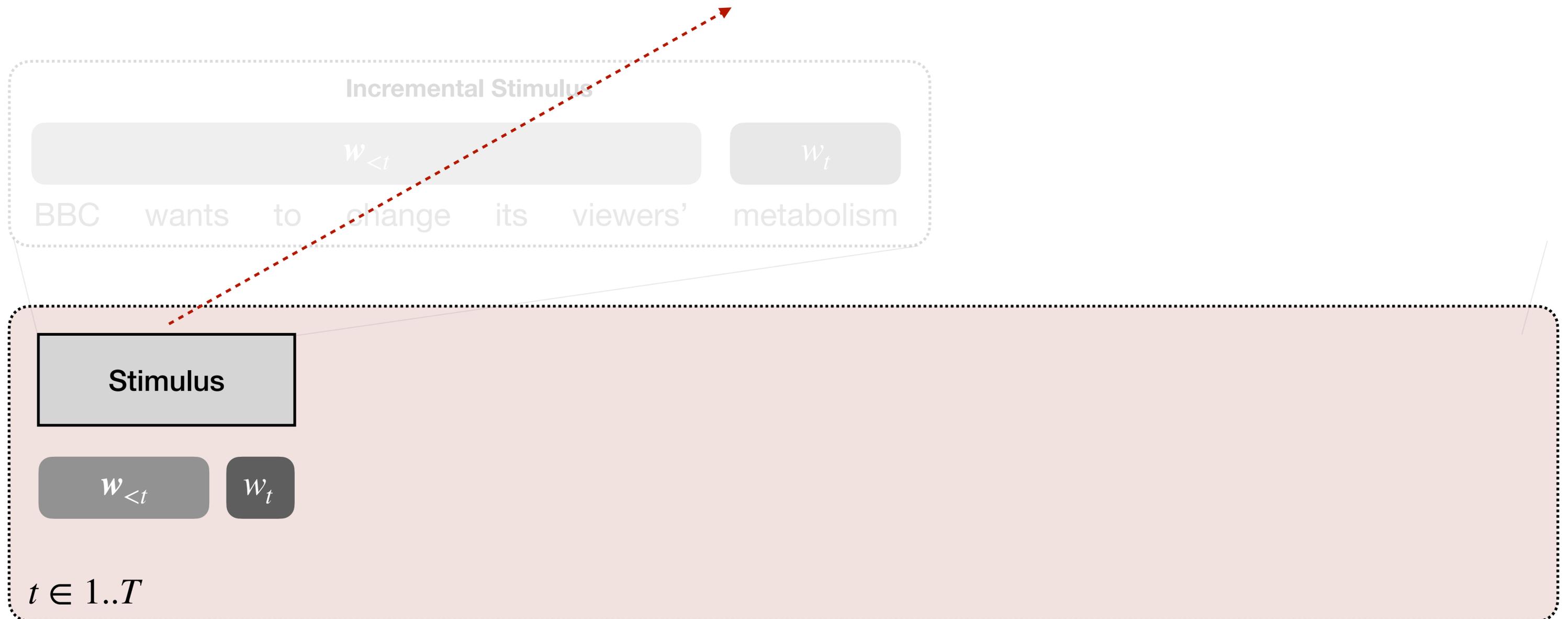


The **Surprisal model** of incremental comprehension



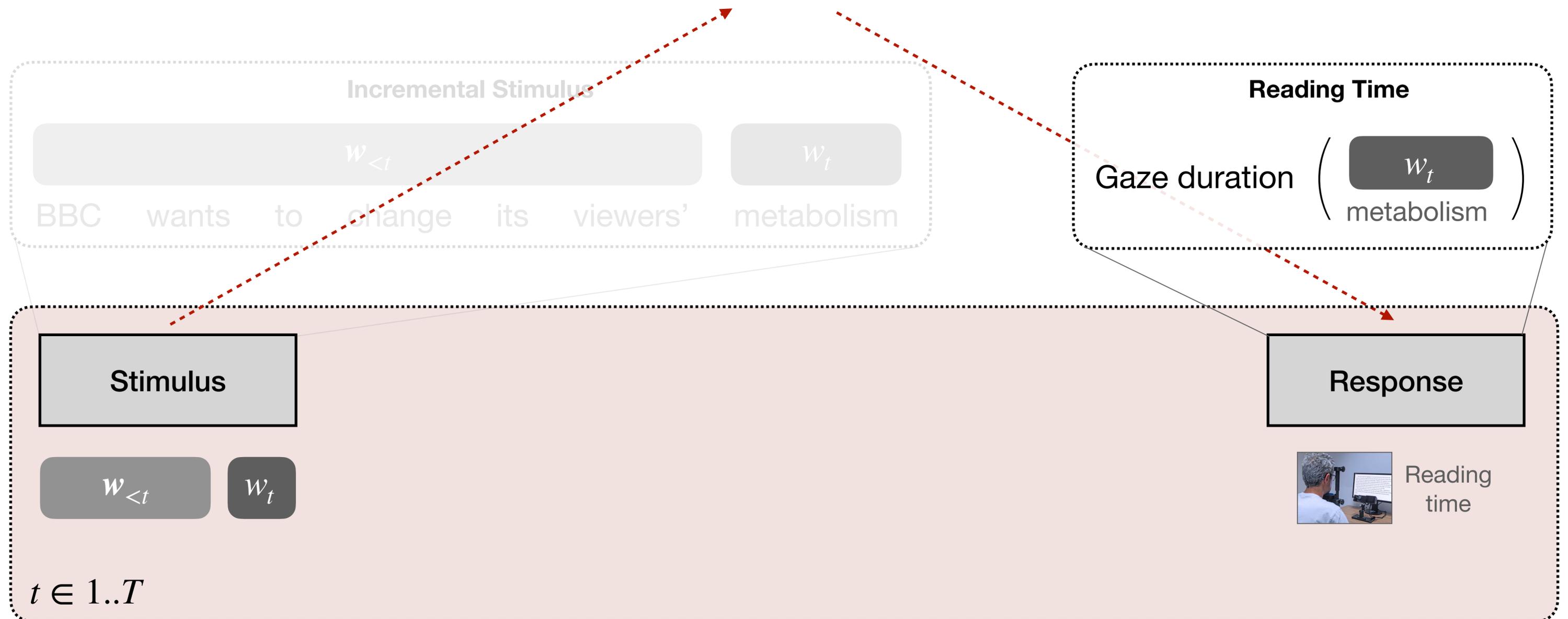
The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



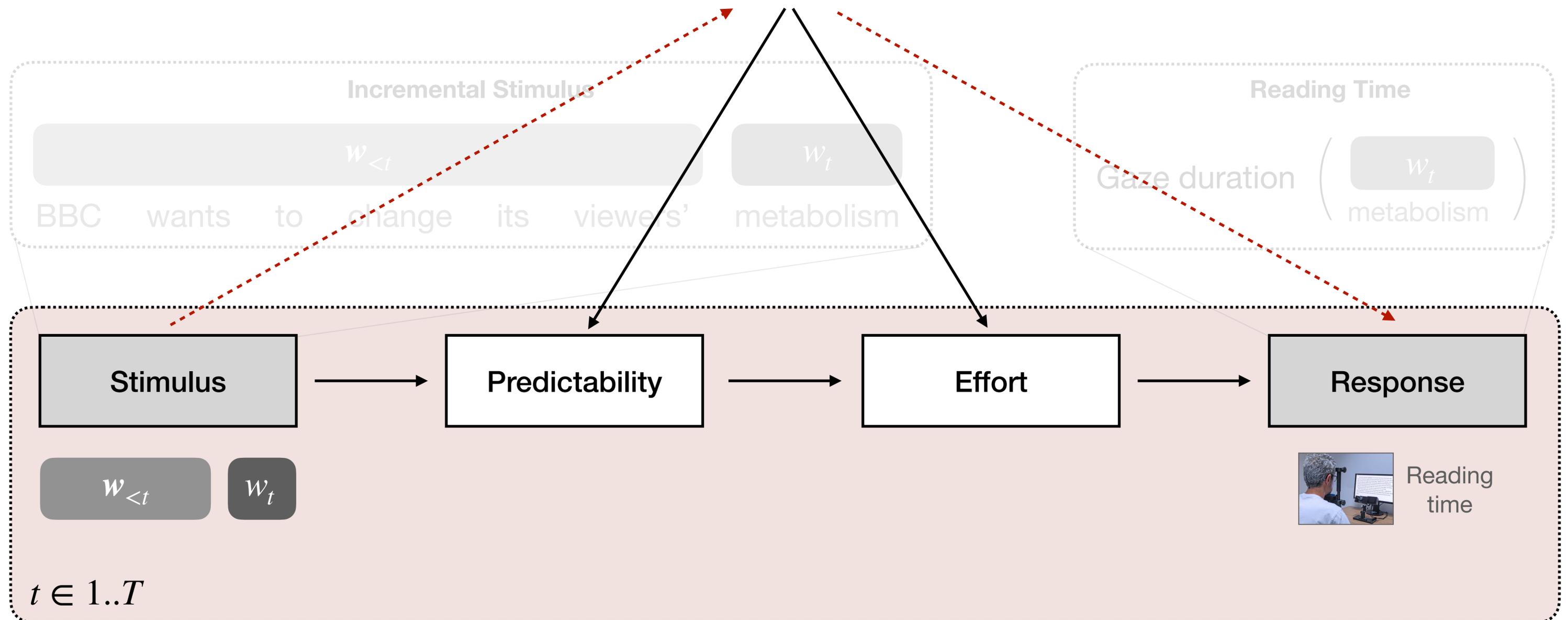
The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



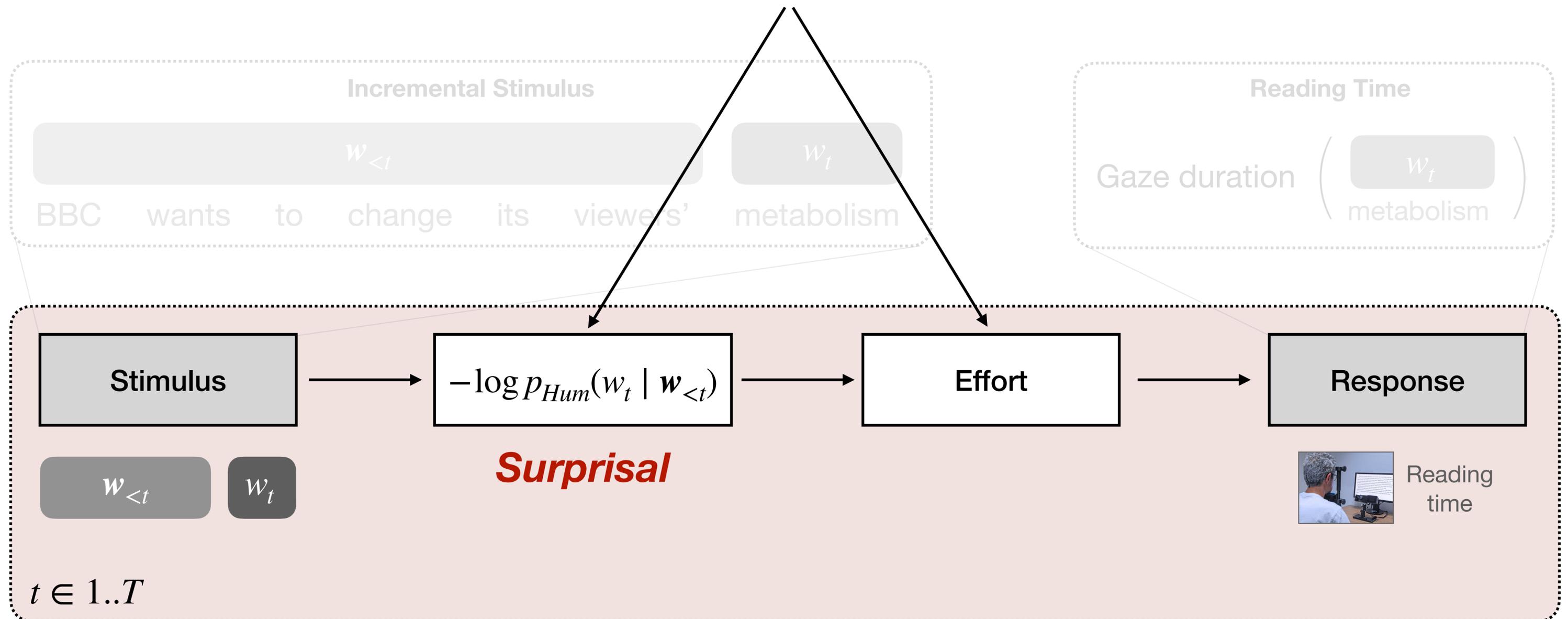
The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



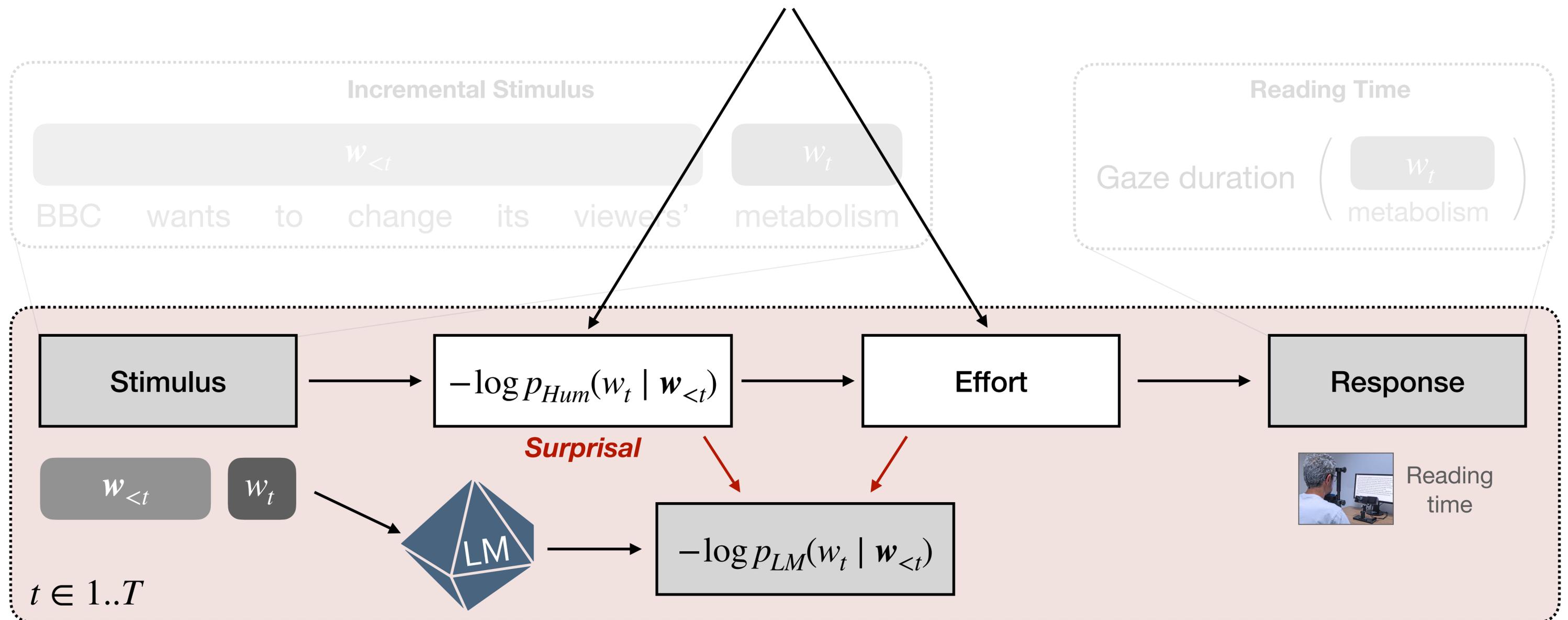
The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



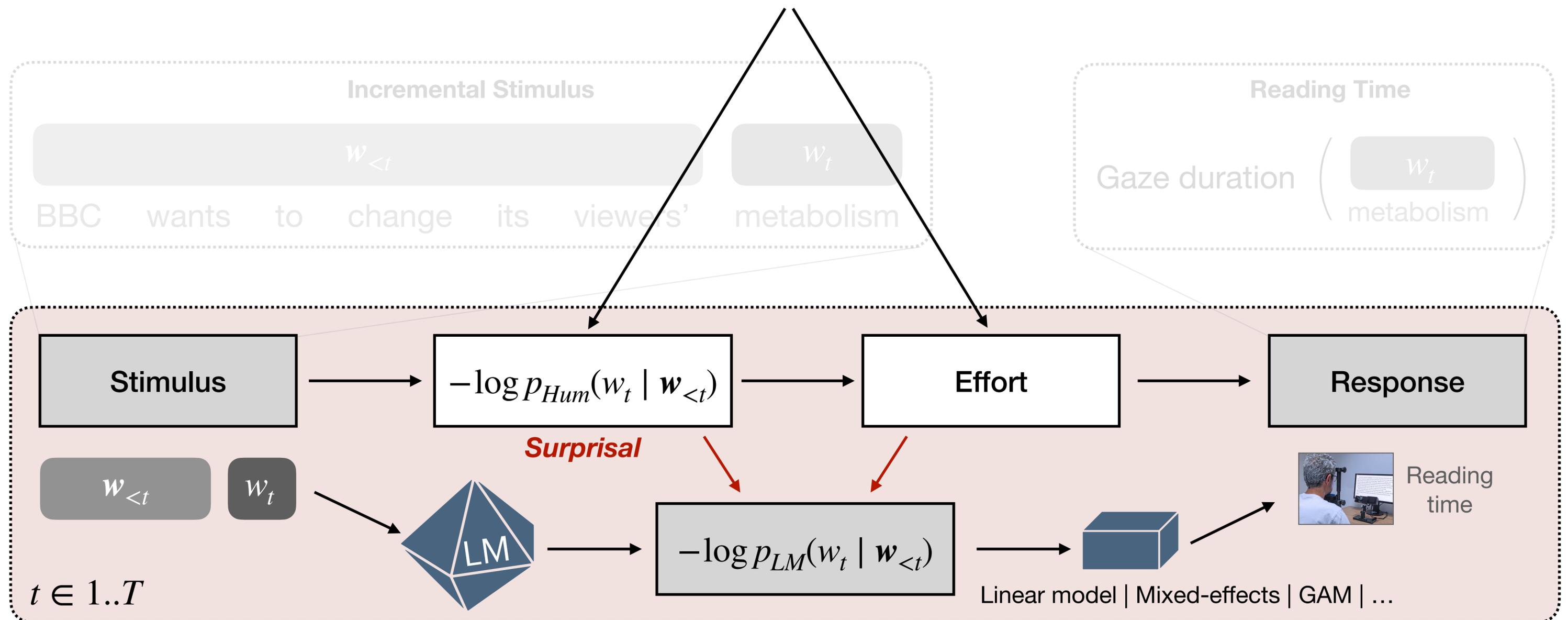
The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



The Surprisal model of incremental comprehension

The Surprisal hypothesis: “The effort a reader must spend to comprehend a word is a function of the word’s contextual predictability.”



Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Data

Dataset: **Multilingual Eye Movement Corpus** (MECO; Siegelman et al., 2022)

- **L1 speakers** (between 29 and 54 per language)
- 12 simplified **Wikipedia-style articles** (1.5-3k words)
- **13 languages** (5 language families)
 - Korean (Koreanic), Turkish (Turkic), Hebrew (Semitic), Finnish (Uralic), Dutch, English, German, Greek, Italian, Russian, and Spanish (Indo-European)

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Data

Dataset: **Multilingual Eye Movement Corpus** (MECO; Siegelman et al., 2022)

- **L1 speakers** (between 29 and 54 per language)
- 12 simplified **Wikipedia-style articles** (1.5-3k words)
- **13 languages** (5 language families)
 - Korean (Koreanic), Turkish (Turkic), Hebrew (Semitic), Finnish (Uralic), Dutch, English, German, Greek, Italian, Russian, and Spanish (Indo-European)
- Predicted variables
 - **First fixation time**: the duration of the first fixation on a word during its first pass.
 - **Gaze duration**: the sum of all first-pass fixations on a word.
 - **Total fixation time**: the sum of all fixations on a word during the trial.

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Data

Dataset: **Multilingual Eye Movement Corpus** (MECO; Siegelman et al., 2022)

- **L1 speakers** (between 29 and 54 per language)
- 12 simplified **Wikipedia-style articles** (1.5-3k words)
- **13 languages** (5 language families)
 - Korean (Koreanic), Turkish (Turkic), Hebrew (Semitic), Finnish (Uralic), Dutch, English, German, Greek, Italian, Russian, and Spanish (Indo-European)
- Predicted variables
 - **First fixation time**: the duration of the first fixation on a word during its first pass.
 - **Gaze duration**: the sum of all first-pass fixations on a word.
 - **Total fixation time**: the sum of all fixations on a word during the trial.

Language models

Monolingual LMs

- decoder-only transformer (6 layers)
- context window size of 512 tokens
- trained on the Wiki40B dataset

Multilingual LM

- pre-trained mGPT
- decoder-only transformer (24 layers)
- trained on Wikipedia + Cleaned Common Crawl Corpus

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Data

Dataset: **Multilingual Eye Movement Corpus** (MECO; Siegelman et al., 2022)

- **L1 speakers** (between 29 and 54 per language)
- 12 simplified **Wikipedia-style articles** (1.5-3k words)
- **13 languages** (5 language families)
 - Korean (Koreanic), Turkish (Turkic), Hebrew (Semitic), Finnish (Uralic), Dutch, English, German, Greek, Italian, Russian, and Spanish (Indo-European)
- Predicted variables
 - **First fixation time**: the duration of the first fixation on a word during its first pass.
 - **Gaze duration**: the sum of all first-pass fixations on a word.
 - **Total fixation time**: the sum of all fixations on a word during the trial.

Language models

Monolingual LMs

- decoder-only transformer (6 layers)
- context window size of 512 tokens
- trained on the Wiki40B dataset

Multilingual LM

- pre-trained mGPT
- decoder-only transformer (24 layers)
- trained on Wikipedia + Cleaned Common Crawl Corpus

Statistical models

Linear regression and generalised additive models (GAMs)

Baseline predictors: word frequency and length

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Baseline predictors

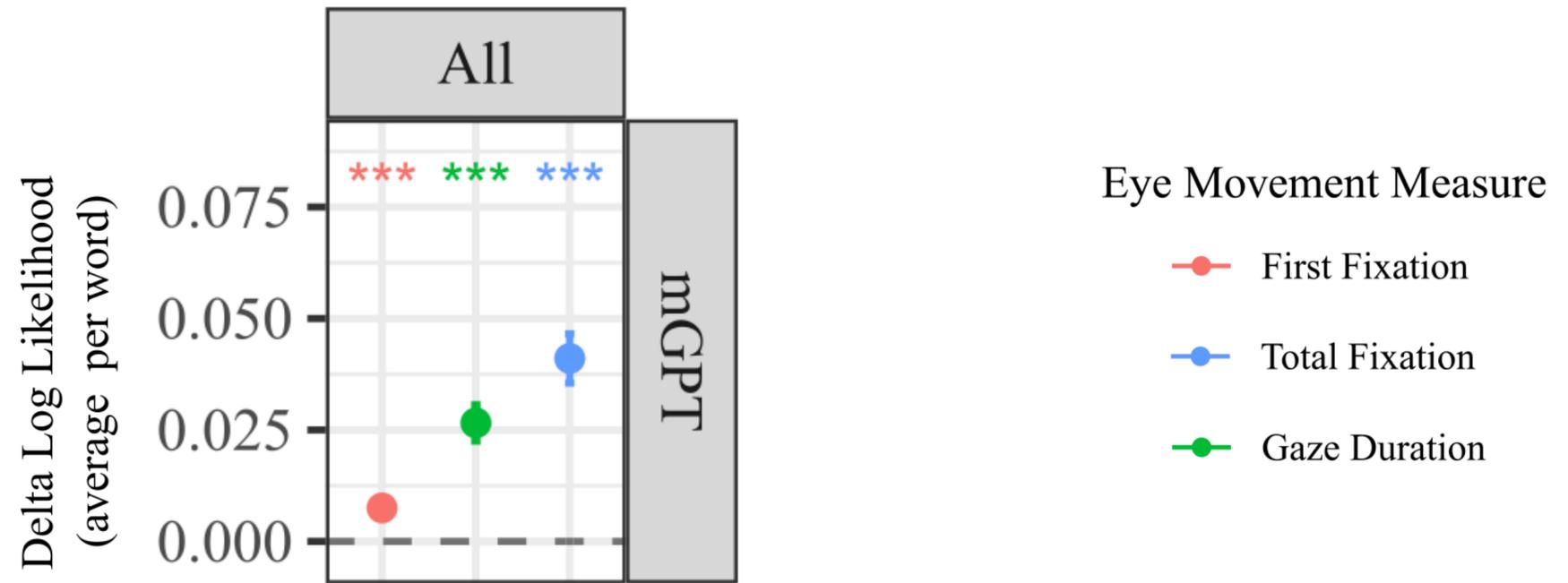
reading time (w_t) \sim surprisal (w_t) + frequency (w_t) + length (w_t)

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Baseline predictors

$$\begin{aligned} \text{reading time } (w_t) \sim & \text{surprisal } (w_t) + \text{frequency } (w_t) + \text{length } (w_t) \\ & + \text{surprisal } (w_{t-1}) + \text{frequency } (w_{t-1}) + \text{length } (w_{t-1}) \\ & + \text{surprisal } (w_{t-2}) + \text{frequency } (w_{t-2}) + \text{length } (w_{t-2}) \end{aligned} \quad \left| \begin{array}{l} \text{Spillover} \\ \text{effect} \end{array} \right.$$

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

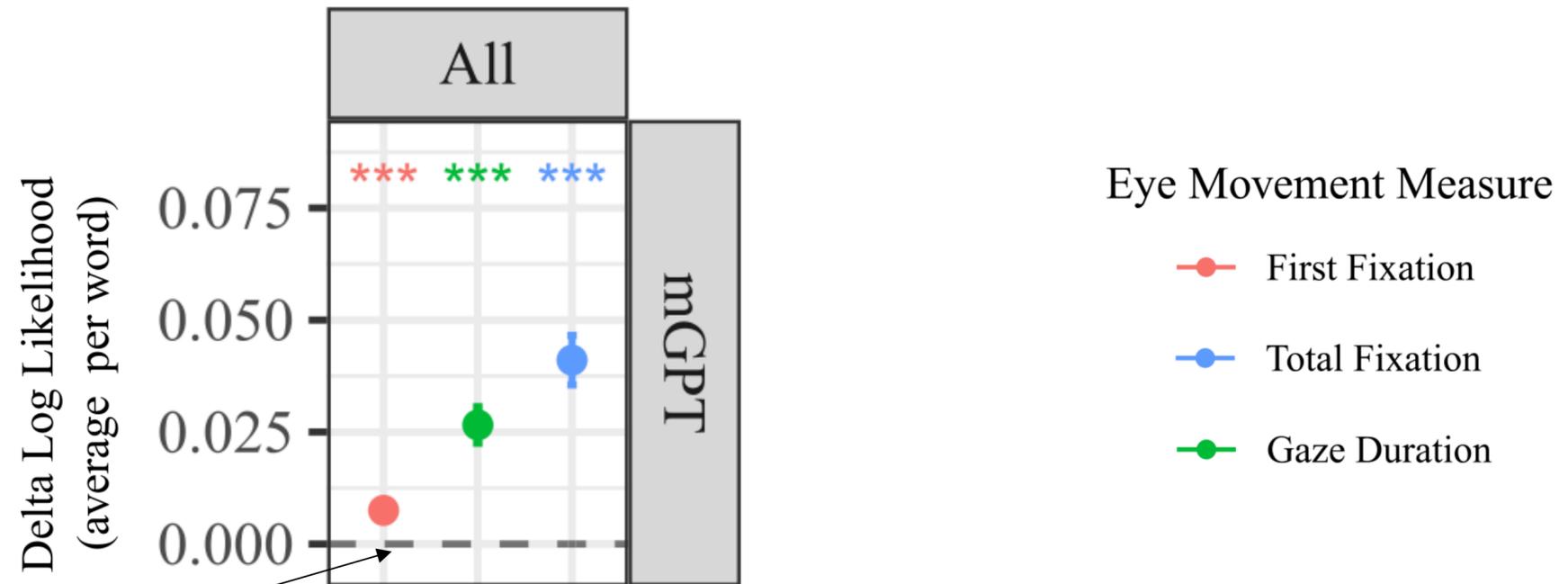


Baseline predictors

$$\begin{aligned}
 \text{reading time } (w_t) \sim & \text{surprisal } (w_t) + \text{frequency } (w_t) + \text{length } (w_t) \\
 & + \text{surprisal } (w_{t-1}) + \text{frequency } (w_{t-1}) + \text{length } (w_{t-1}) \\
 & + \text{surprisal } (w_{t-2}) + \text{frequency } (w_{t-2}) + \text{length } (w_{t-2})
 \end{aligned}$$

Spillover effect

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)



Baseline:

$$\text{reading time } (w_t) \sim \text{frequency } (w_t) + \text{length } (w_t)$$

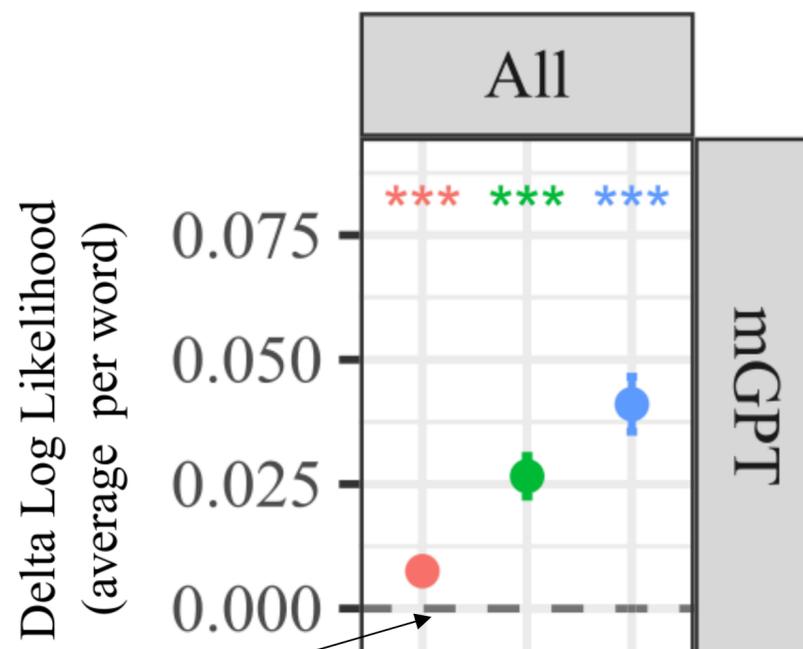
Baseline predictors

$$\begin{aligned} \text{reading time } (w_t) \sim & \text{surprisal } (w_t) + \text{frequency } (w_t) + \text{length } (w_t) \\ & + \text{surprisal } (w_{t-1}) + \text{frequency } (w_{t-1}) + \text{length } (w_{t-1}) \\ & + \text{surprisal } (w_{t-2}) + \text{frequency } (w_{t-2}) + \text{length } (w_{t-2}) \end{aligned}$$

Spillover effect

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

Predictive power
How much better than
the baseline is the
surprisal model?



Eye Movement Measure

- First Fixation
- Total Fixation
- Gaze Duration

Baseline:

$$\text{reading time } (w_t) \sim \text{frequency } (w_t) + \text{length } (w_t)$$

Baseline predictors

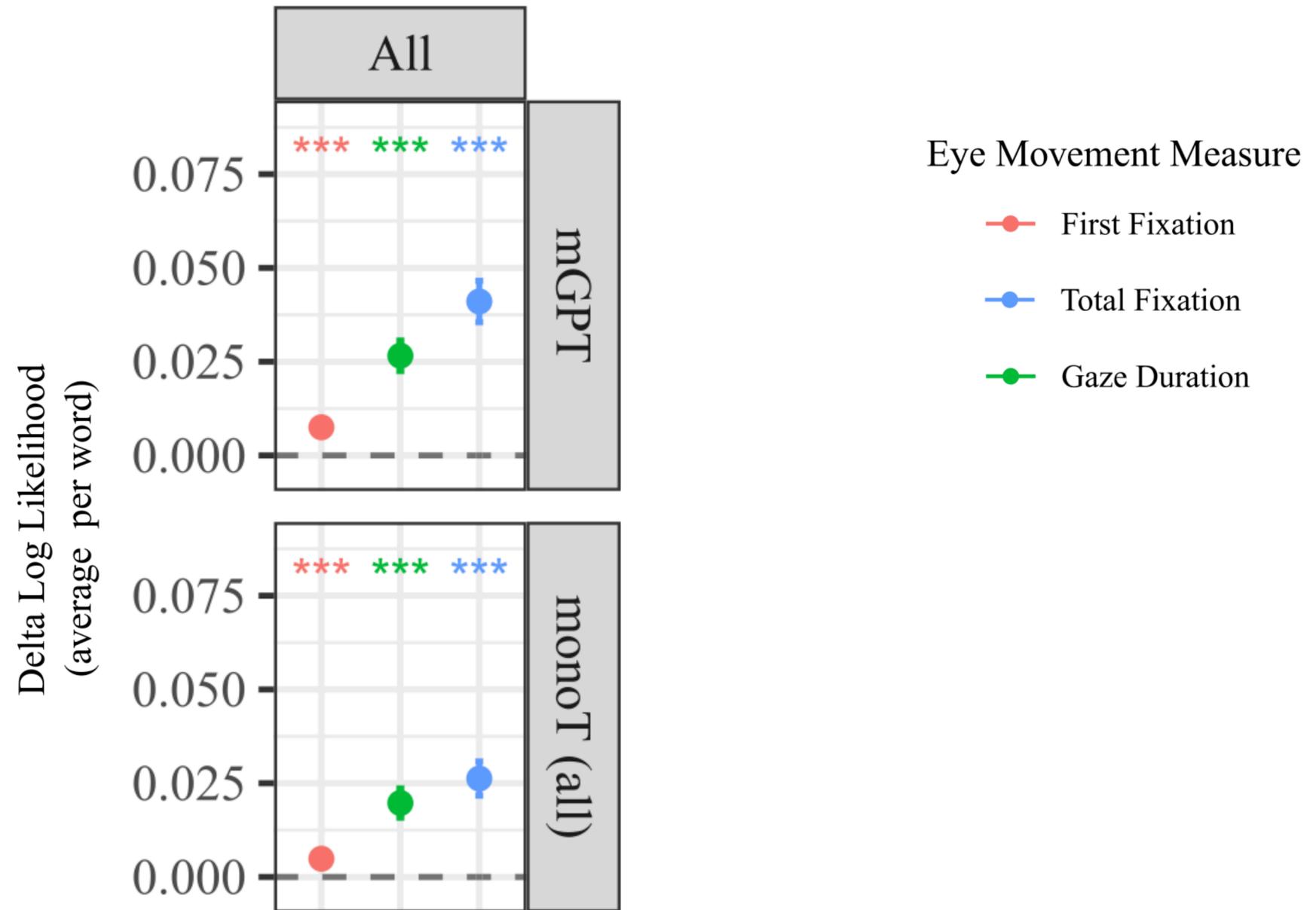


$$\begin{aligned} \text{reading time } (w_t) \sim & \text{surprisal } (w_t) + \text{frequency } (w_t) + \text{length } (w_t) \\ & + \text{surprisal } (w_{t-1}) + \text{frequency } (w_{t-1}) + \text{length } (w_{t-1}) \\ & + \text{surprisal } (w_{t-2}) + \text{frequency } (w_{t-2}) + \text{length } (w_{t-2}) \end{aligned}$$

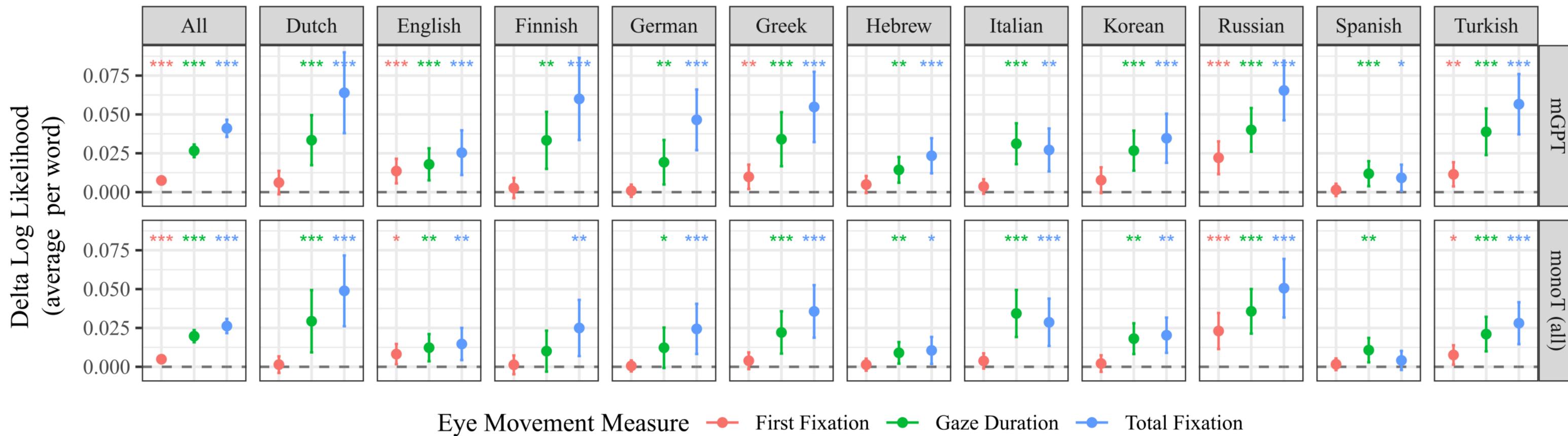
Spillover
effect

Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

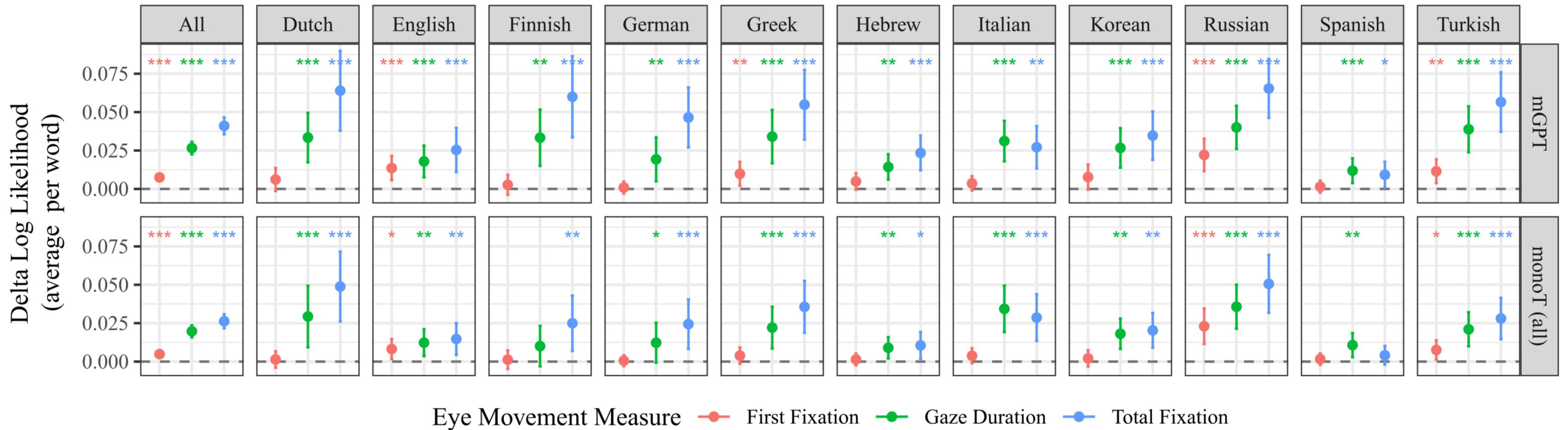
Predictive power
How much better than
the baseline is the
surprisal model?



Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)



Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

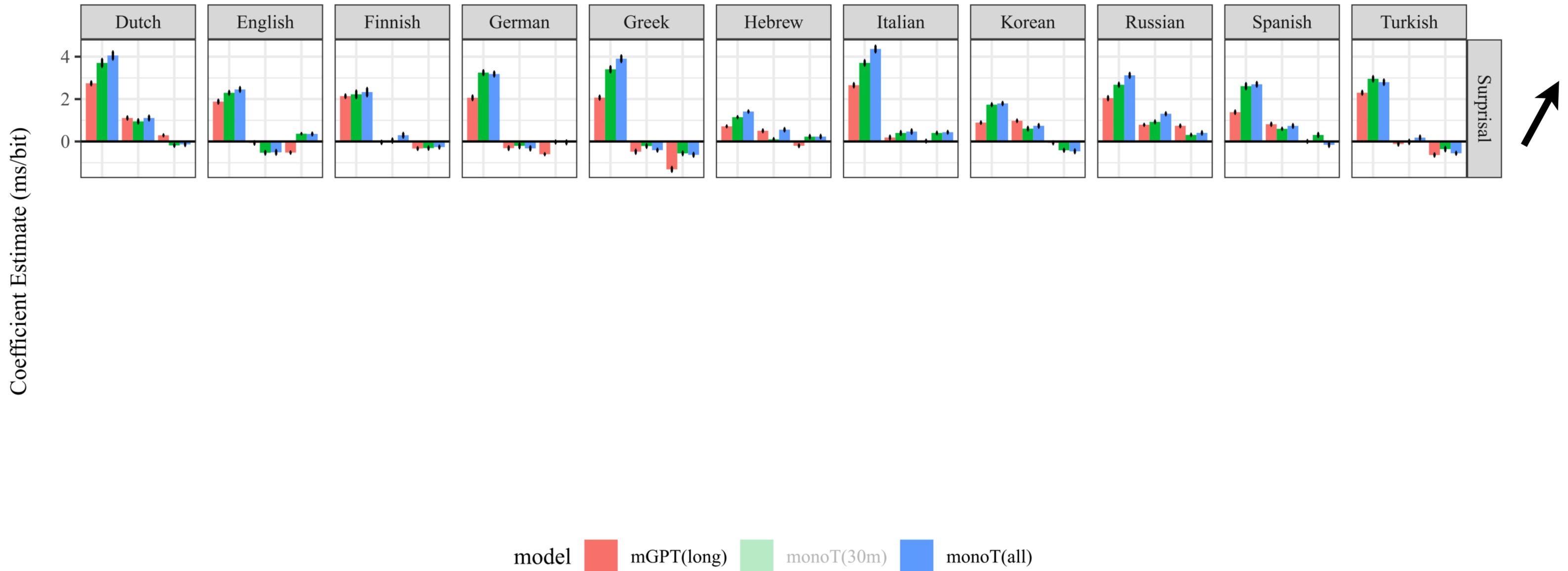


Surprisal predicts reading time above and beyond control predictors

Predictive power: first fixation < gaze duration < total fixation

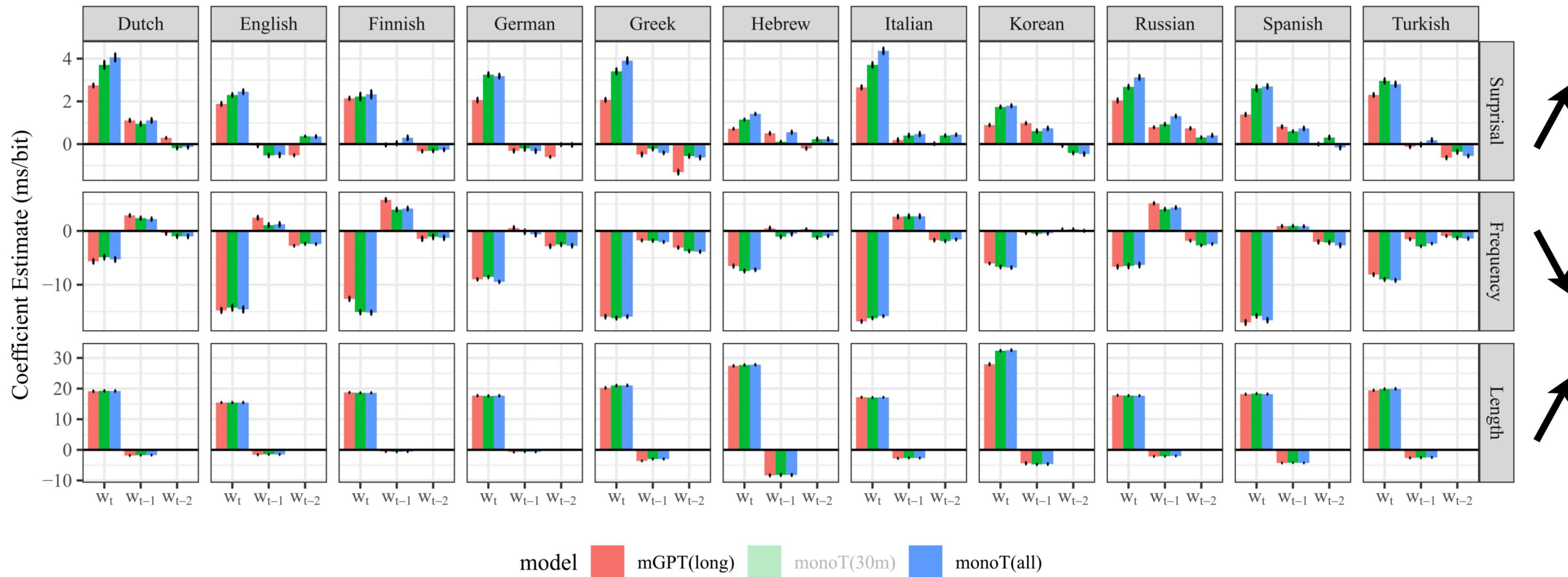
Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

What is the relationship between gaze duration and its predictors?



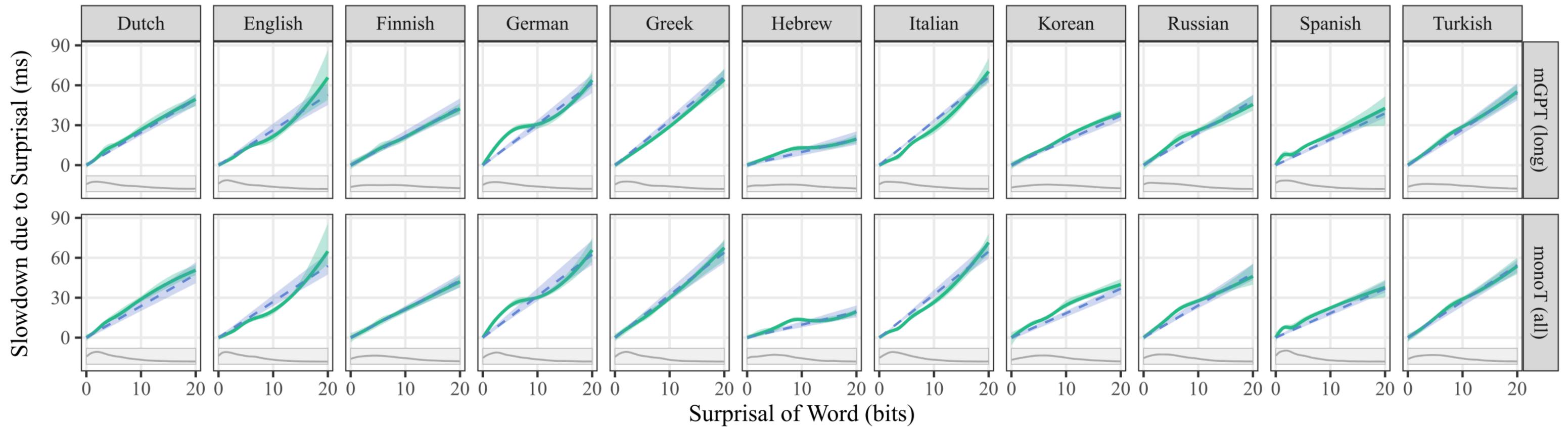
Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

What is the relationship between gaze duration and its predictors?



Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

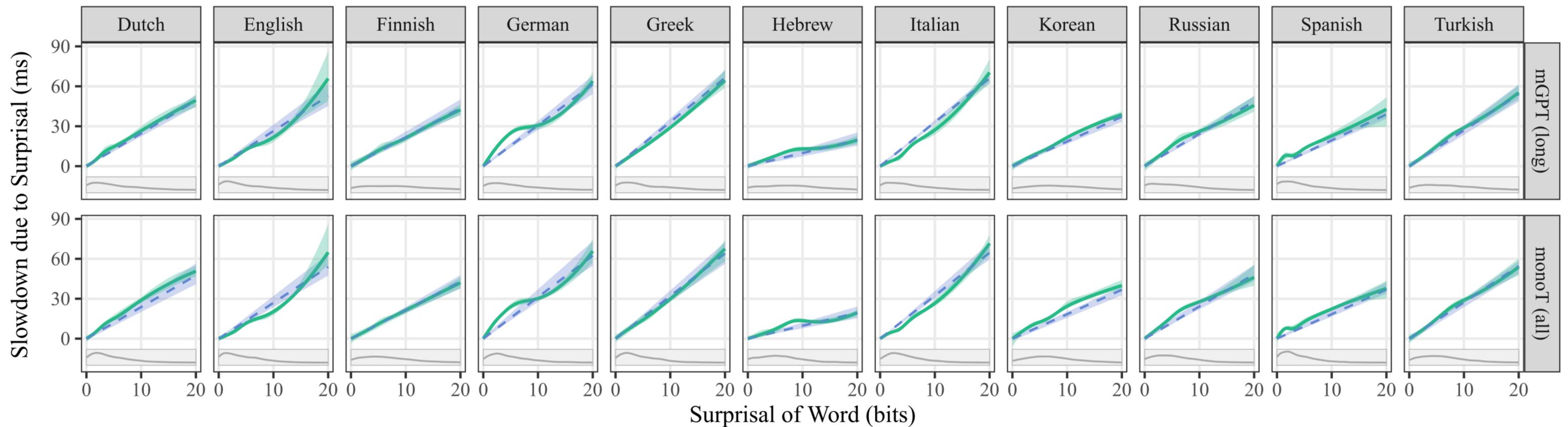
What is the relationship between gaze duration and its predictors?



Generalise additive models (GAMs)

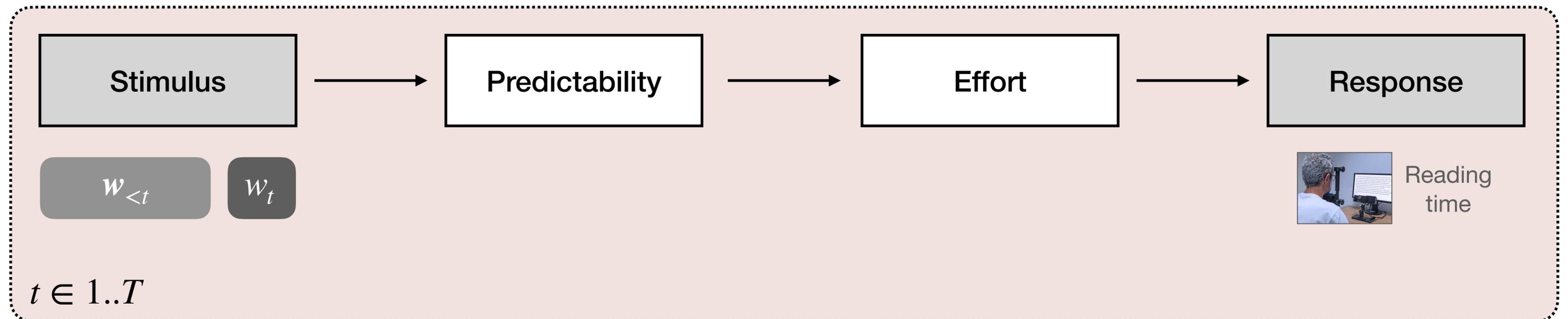
Testing the Predictions of Surprisal Theory in 11 Languages (Wilcox et al. 2023, *TACL*)

What is the relationship between gaze duration and its predictors?

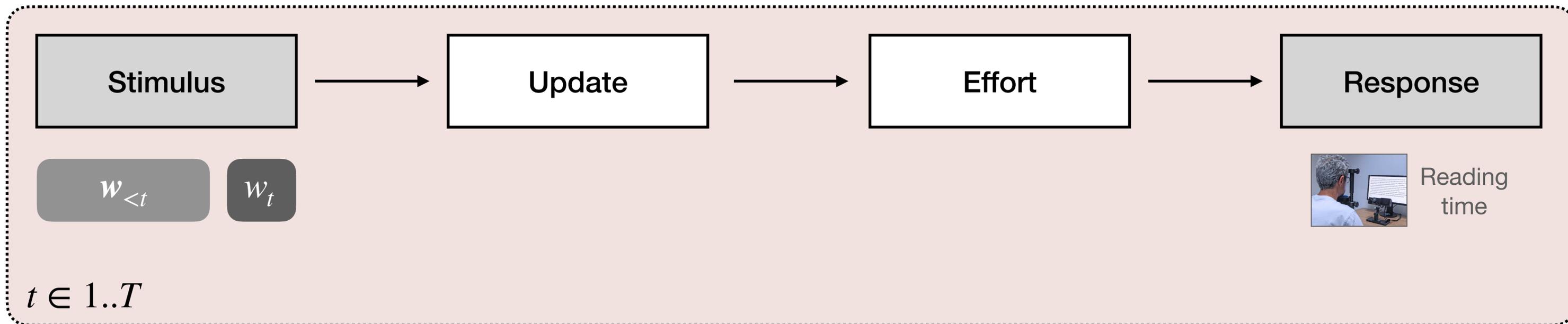
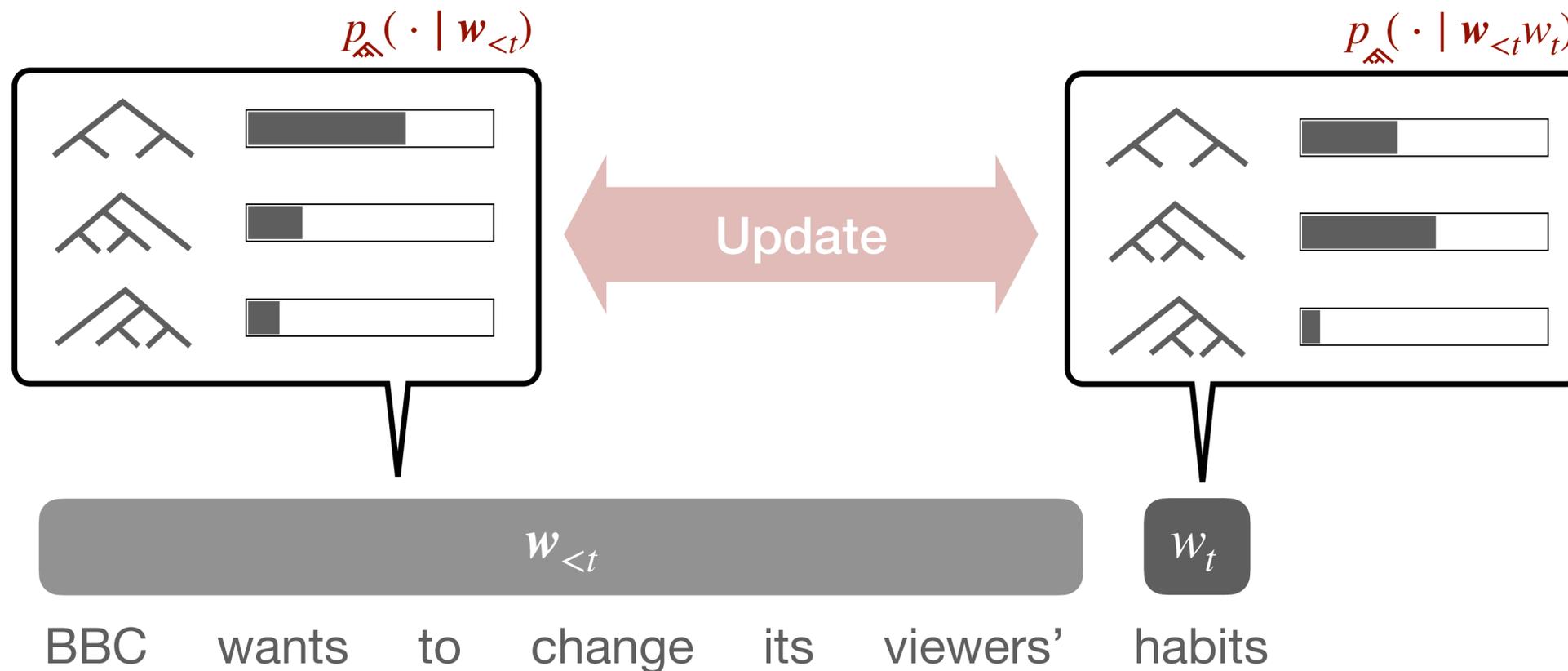


Linear relationship between surprisal and reading time
(= Logarithmic relationship between probability and reading time)

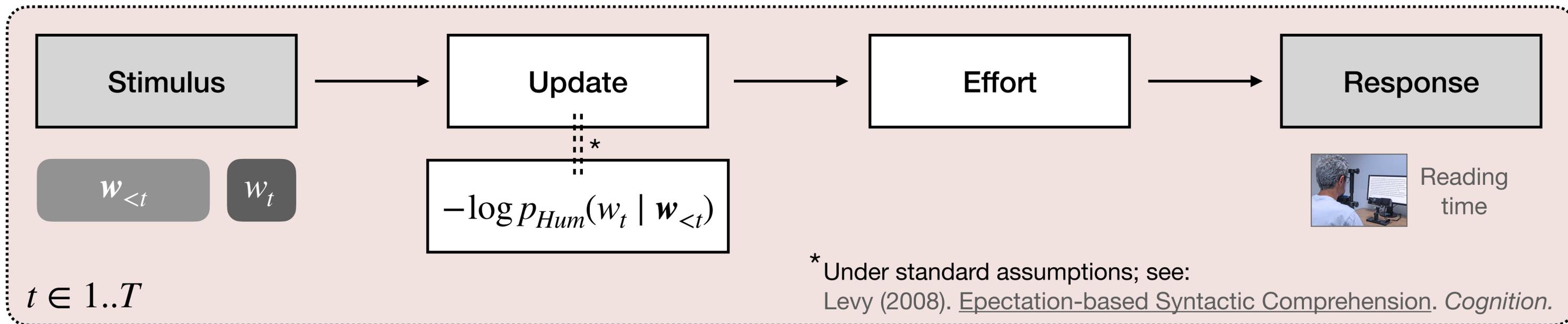
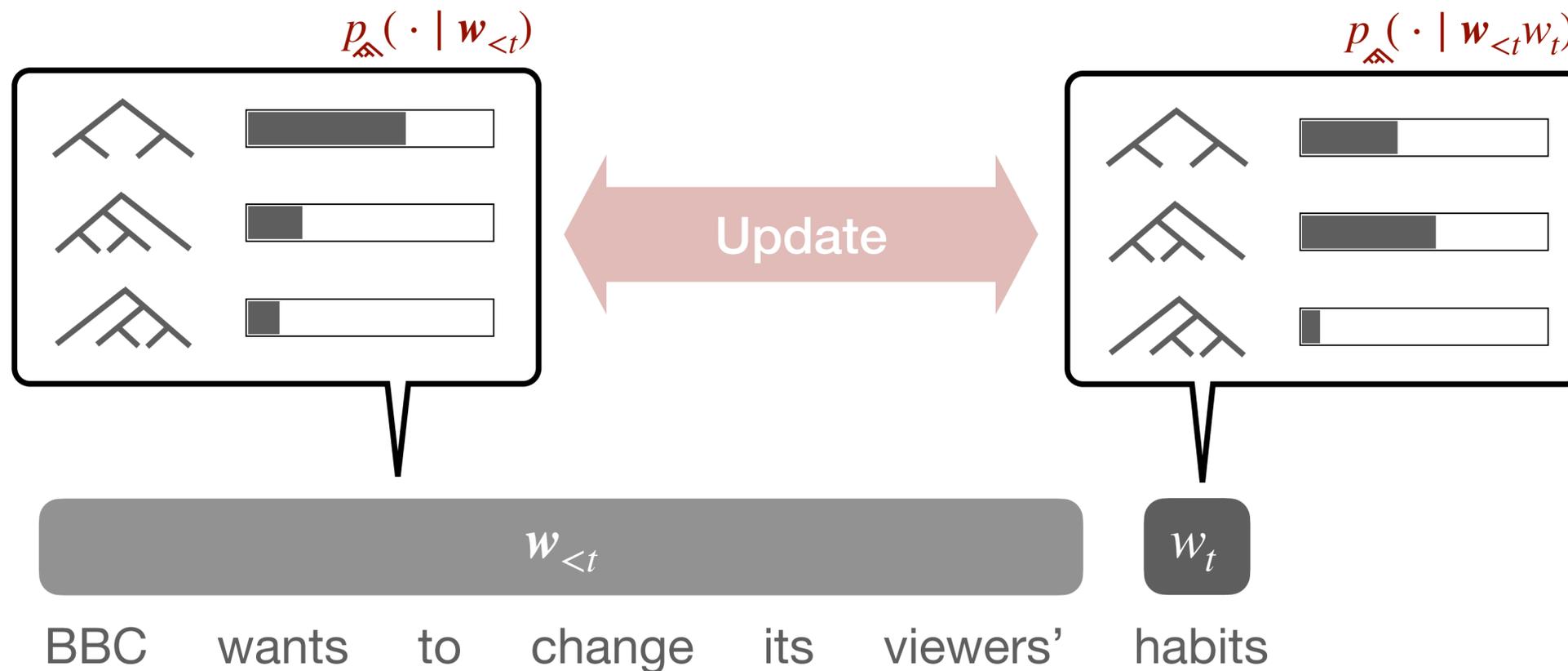
Interpretations of Surprisal



Interpretations of Surprisal



Interpretations of Surprisal



* Under standard assumptions; see:
 Levy (2008). *Expectation-based Syntactic Comprehension*. *Cognition*.

Interpretations of Surprisal

$w_{<t}$
BBC wants to change its viewers'

political

mind

habits

life

customs

Stimulus

Prediction

Response

$w_{<t}$

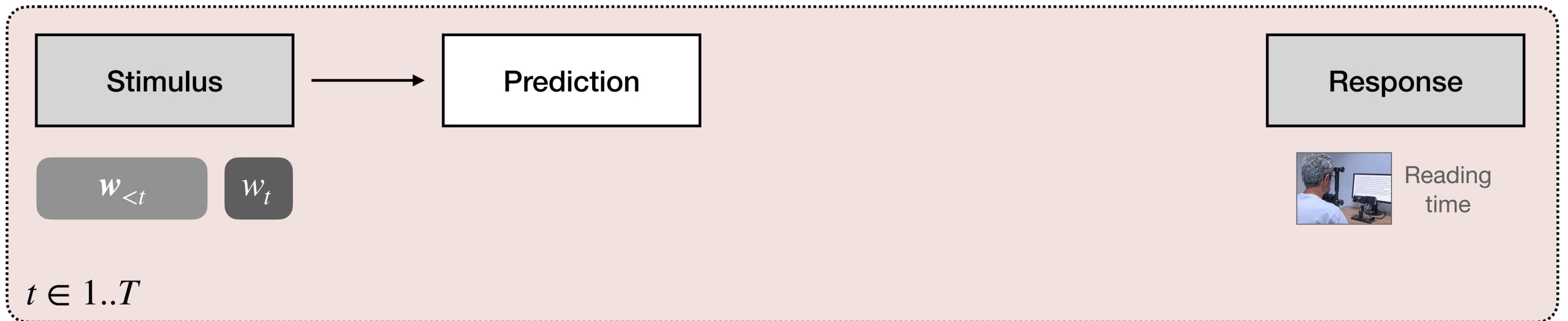
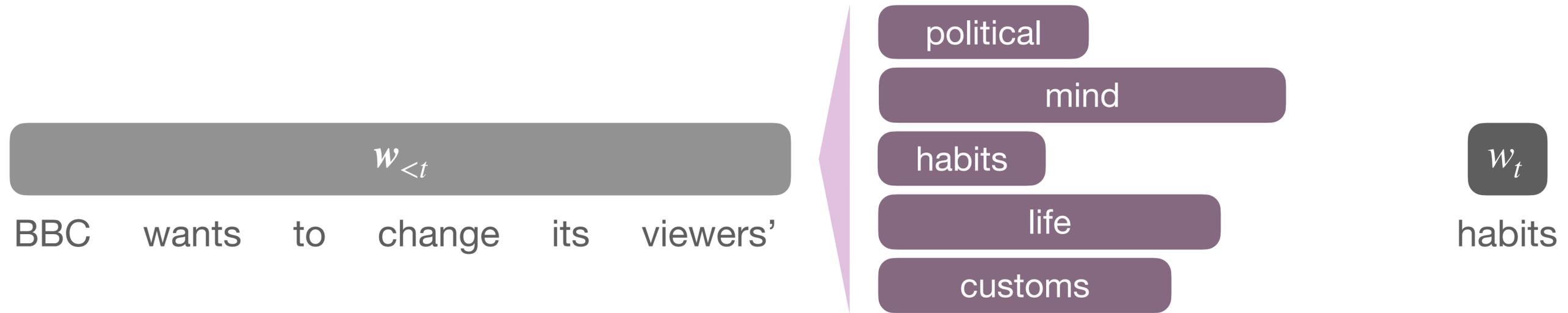
w_t



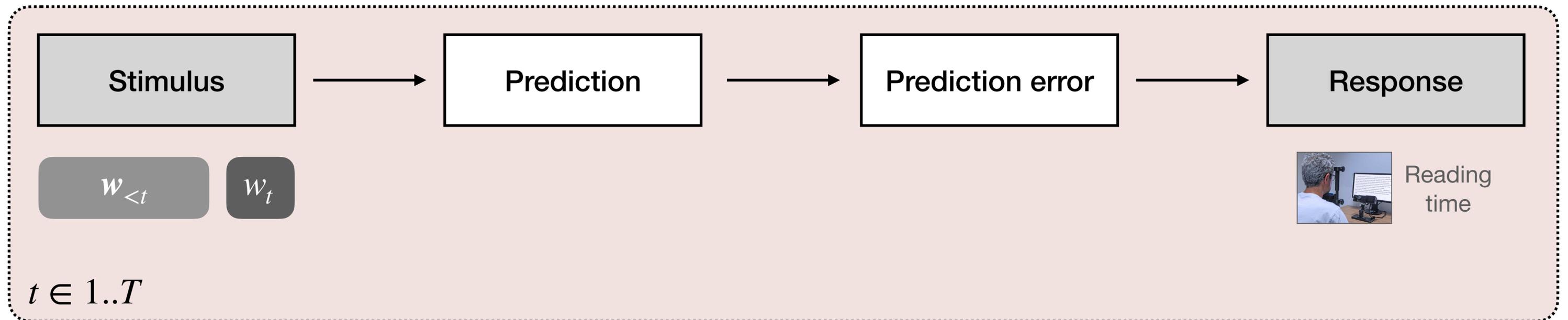
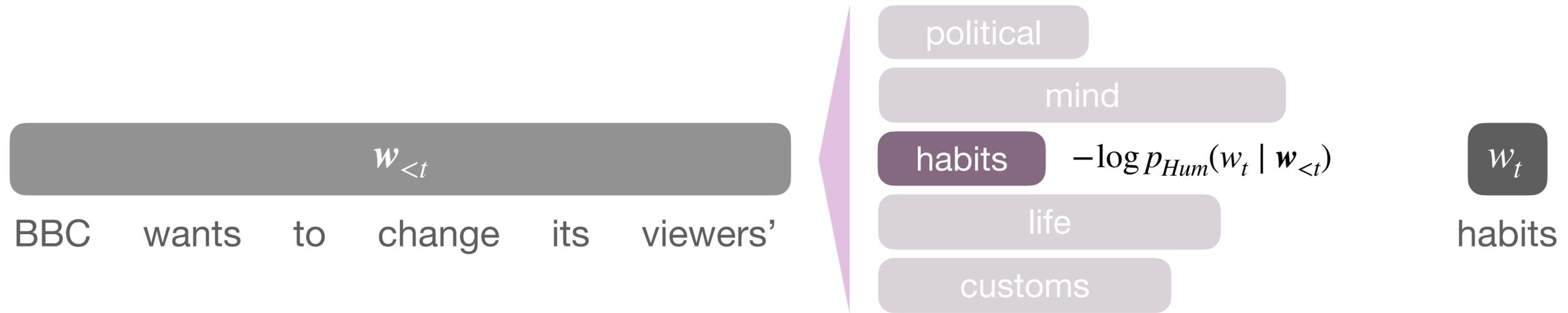
Reading time

$t \in 1..T$

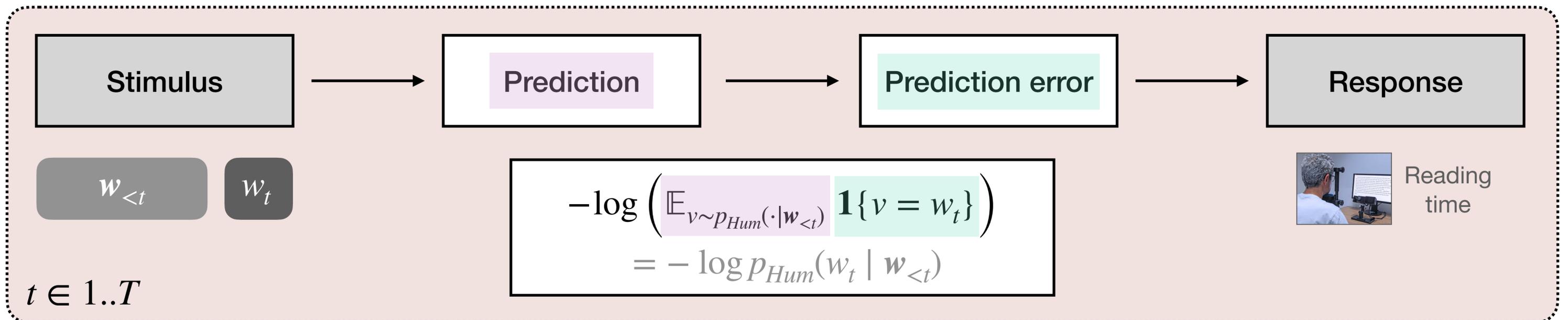
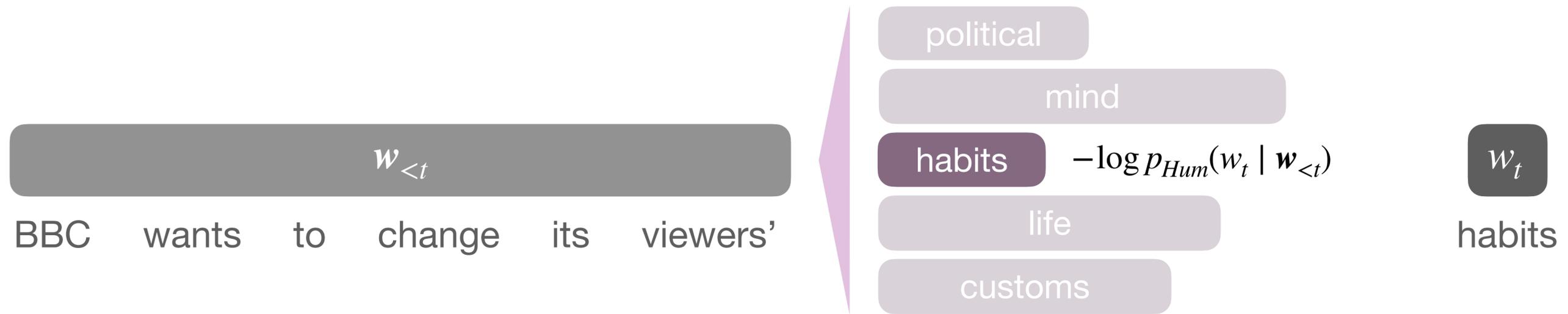
Interpretations of Surprisal



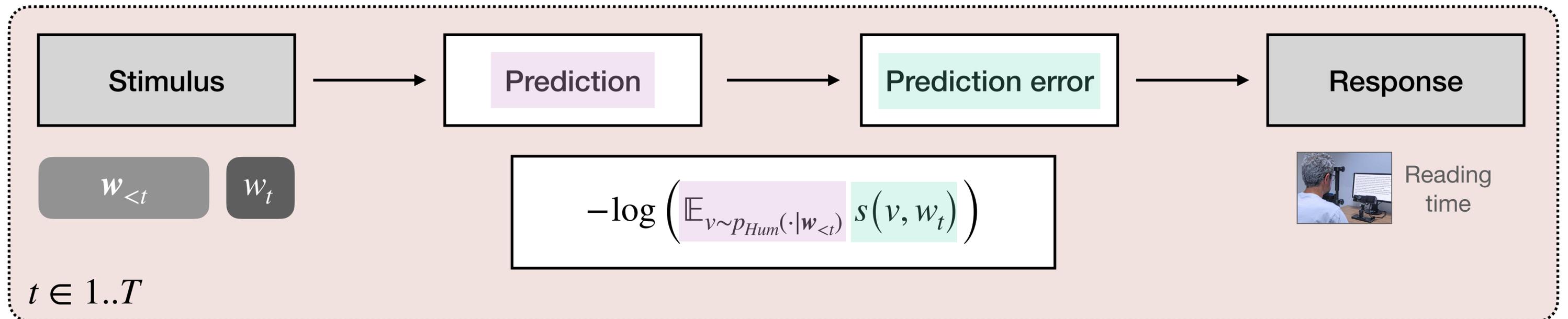
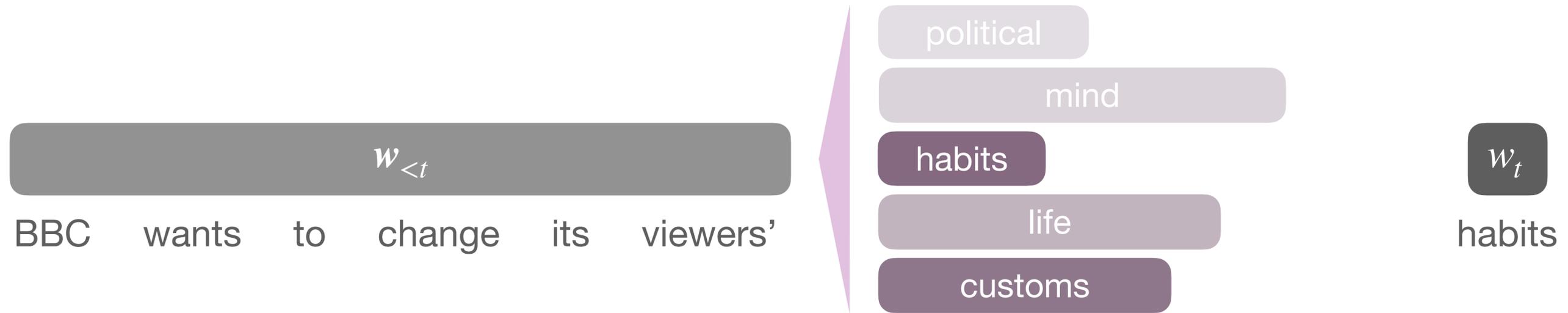
Interpretations of Surprisal



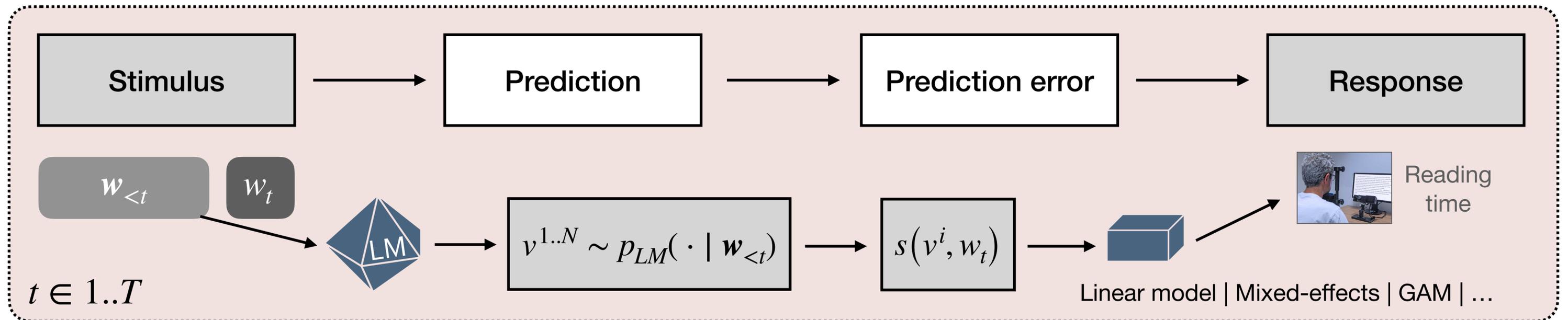
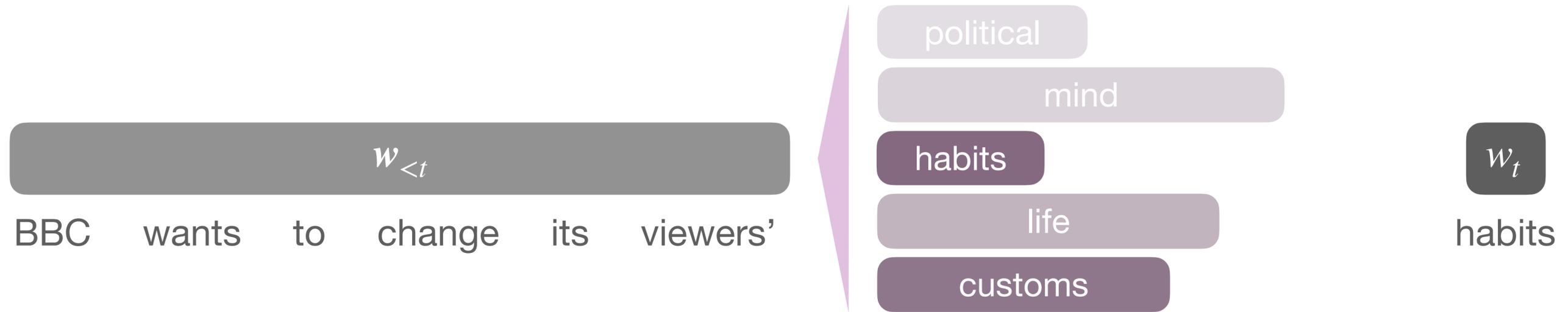
Interpretations of Surprisal



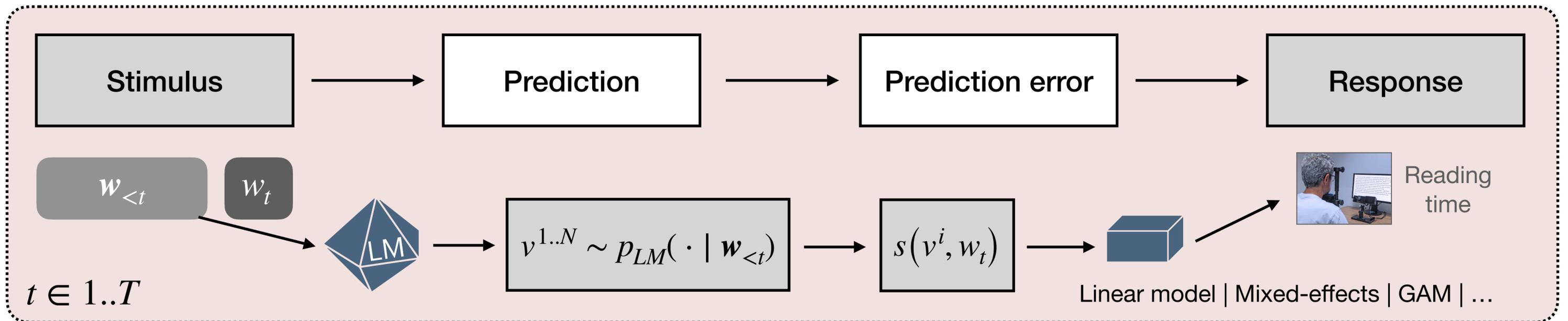
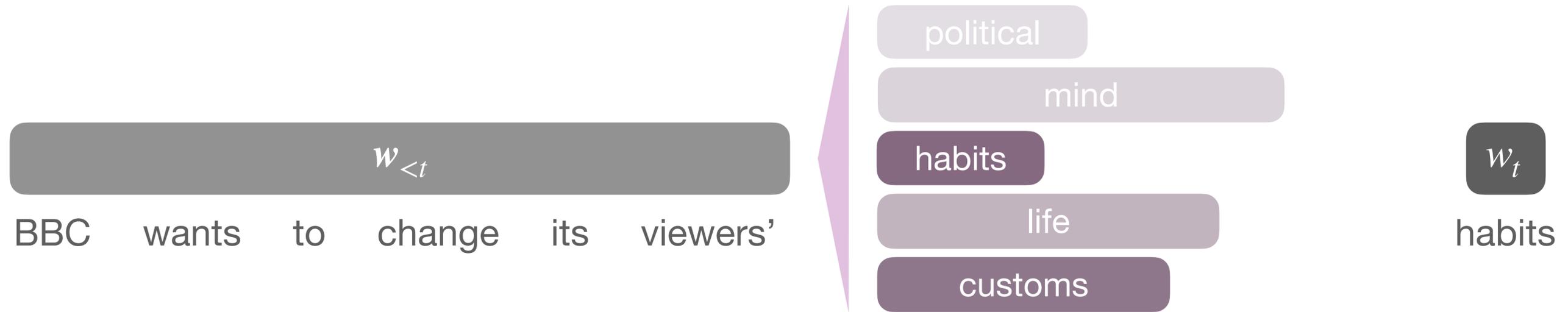
Beyond Surprisal



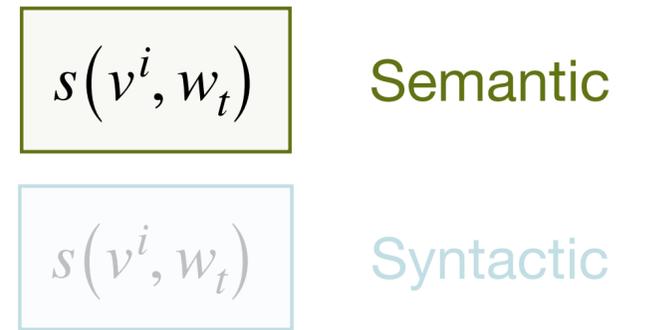
Beyond Surprisal



Beyond Surprisal



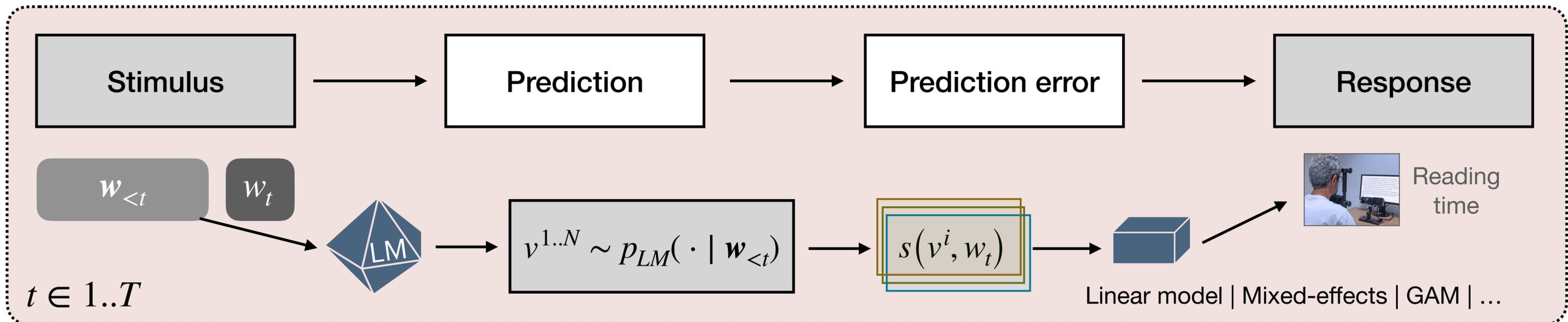
Beyond Surprisal



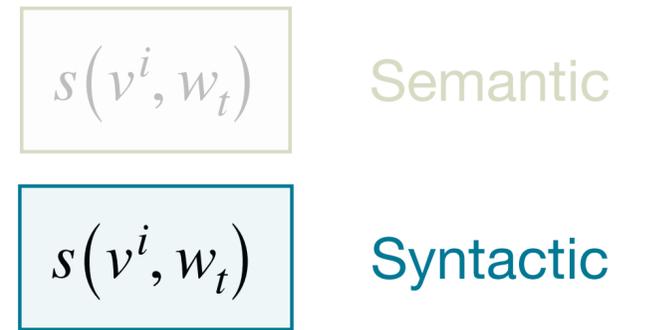
$w_{<t}$
 BBC wants to change its viewers'



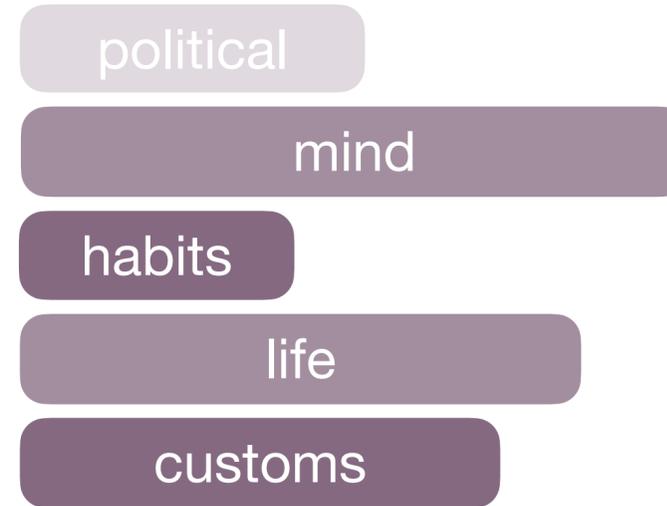
w_t
 diapers



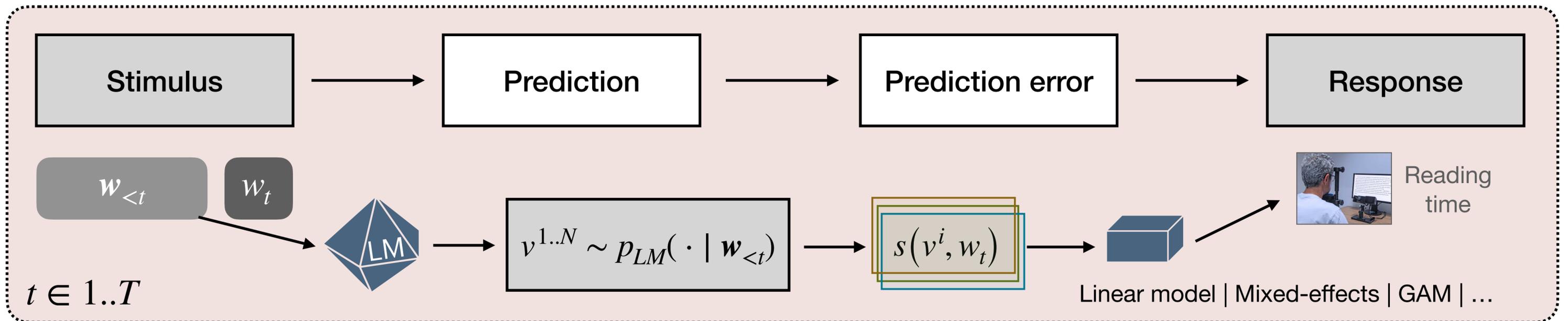
Beyond Surprisal



$w_{<t}$
 BBC wants to change its viewers'



w_t
 diapers



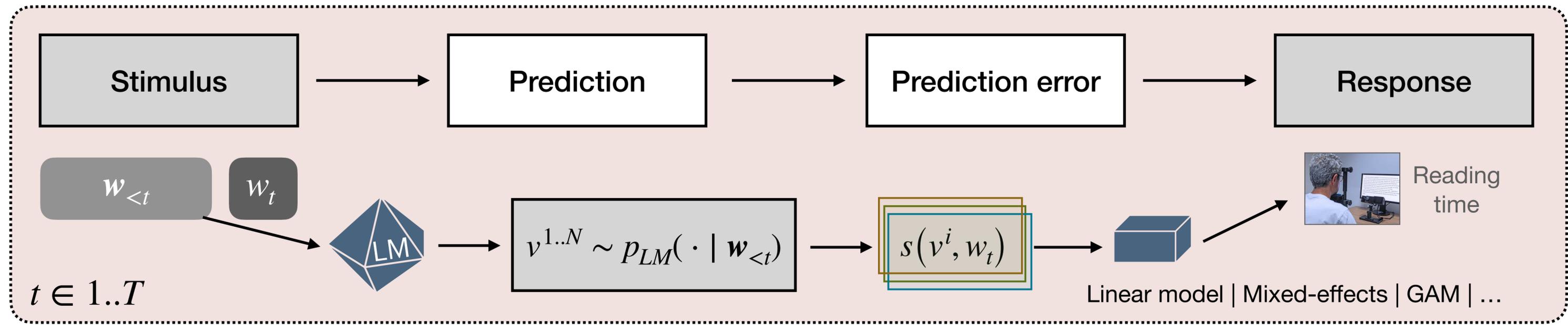
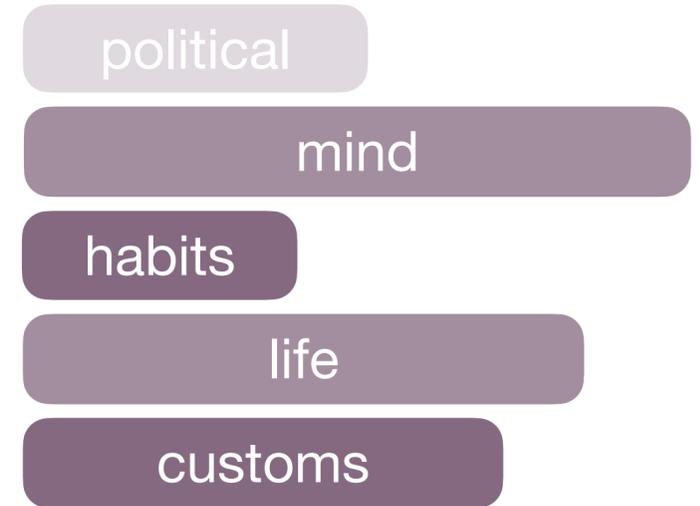
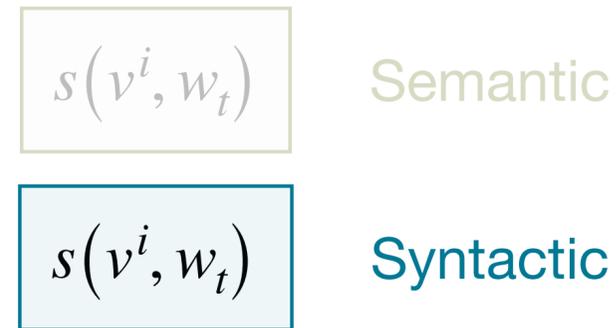
Giulianelli et al. (2023). Information Value: Measuring Utterance Predictability as Distance from Plausible Alternatives. *EMNLP*.

Giulianelli et al. (2024). Generalized Measures of Anticipation and Responsivity in Online Language Processing. *ACL Findings*.

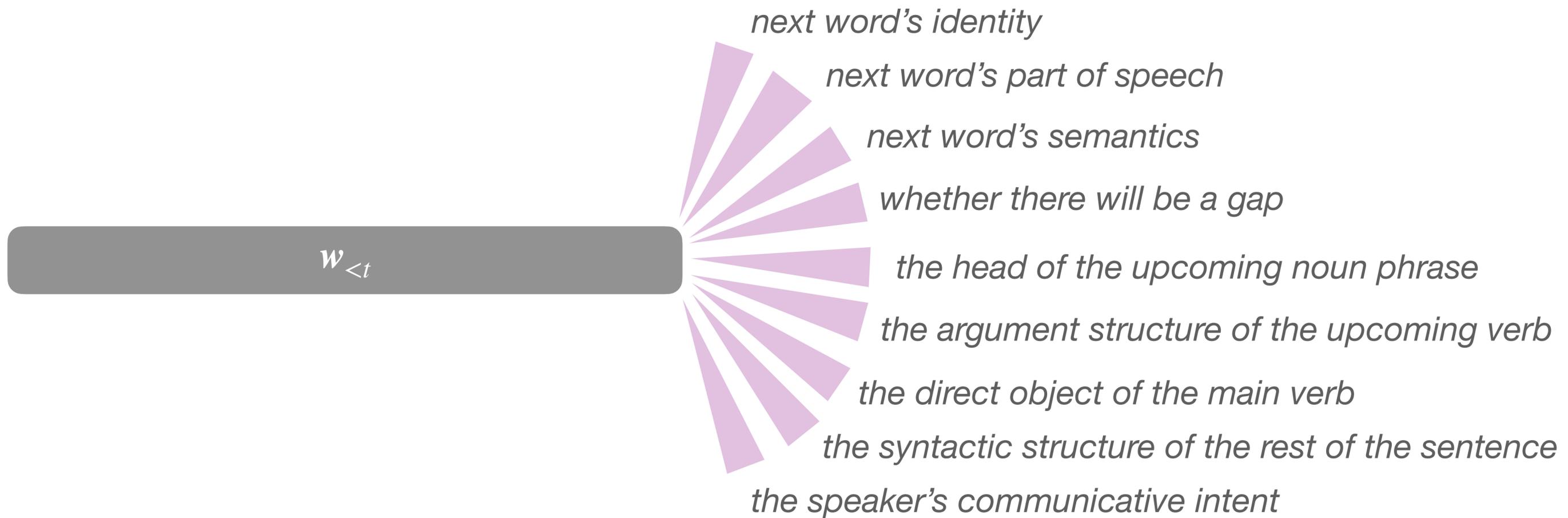
Meister et al. (2024). Towards a Similarity-adjusted Surprisal Theory. *EMNLP*.

Giulianelli et al. (2026). Incremental alternative sampling as a lens into the temporal and representational resolution of linguistic prediction. *JML*.

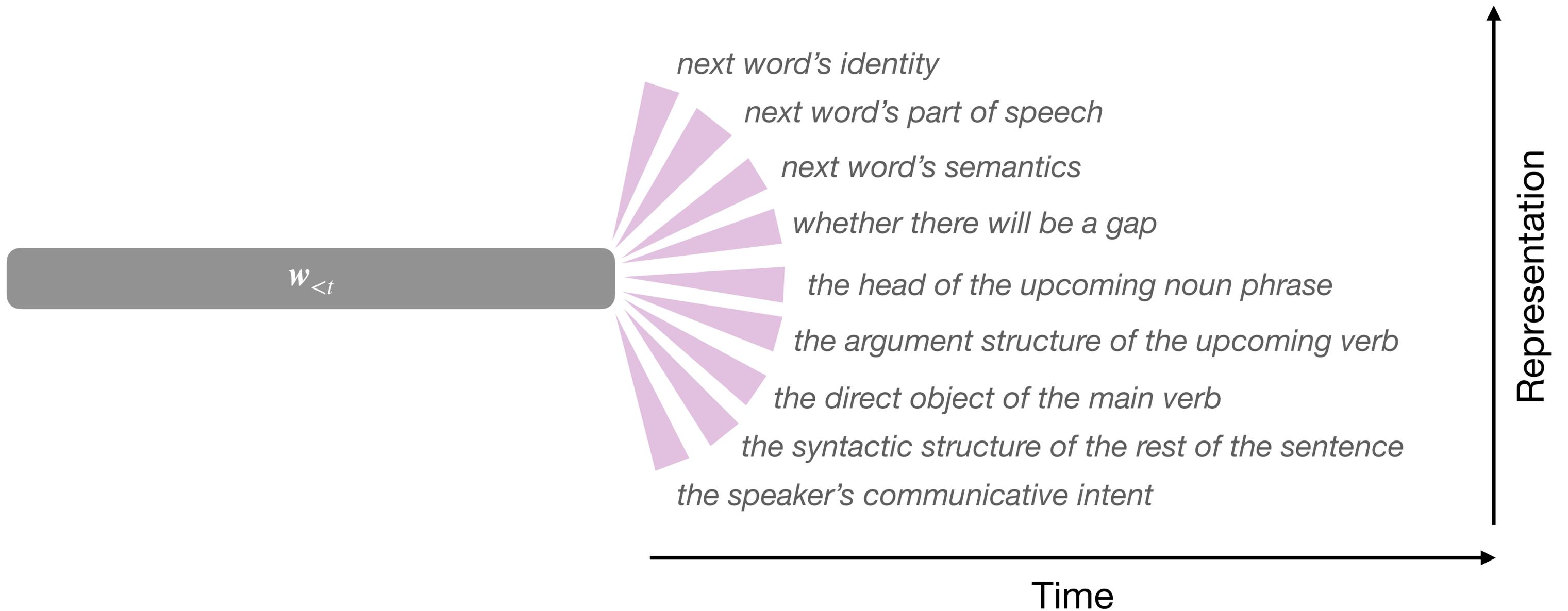
Beyond Surprisal



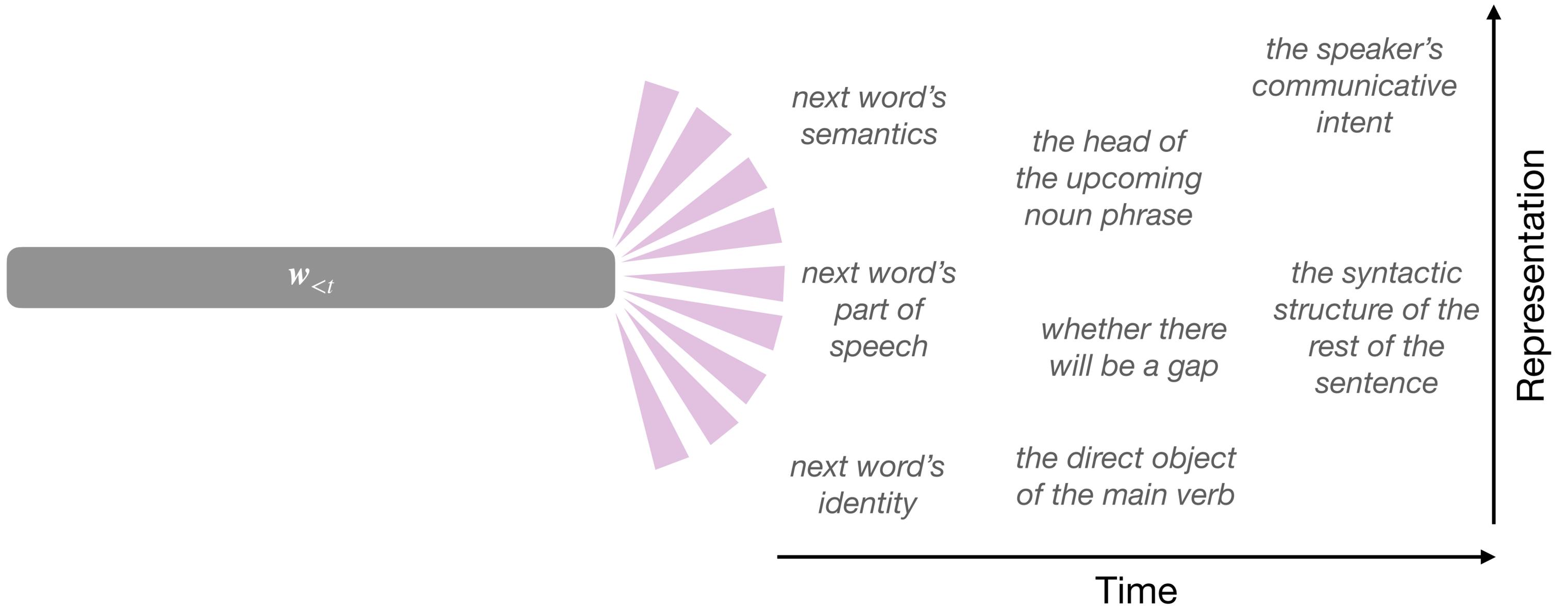
The dimensions of prediction



The dimensions of prediction

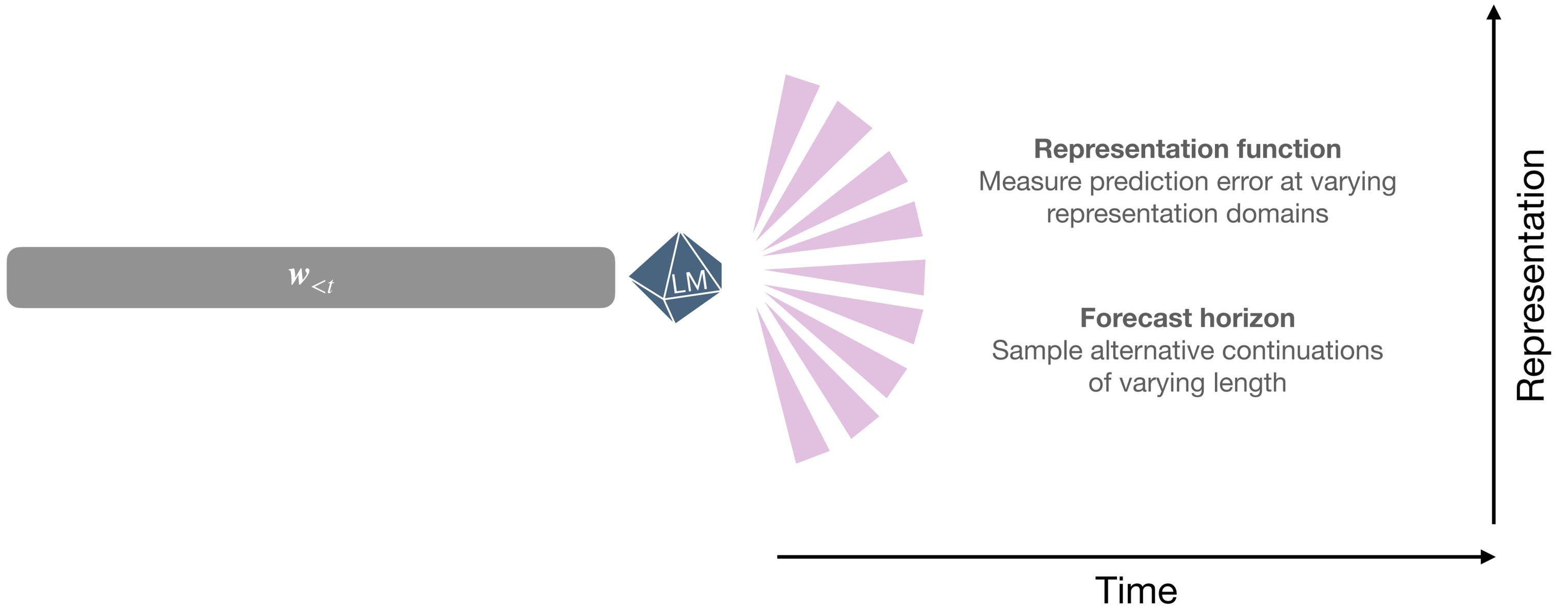


The dimensions of prediction

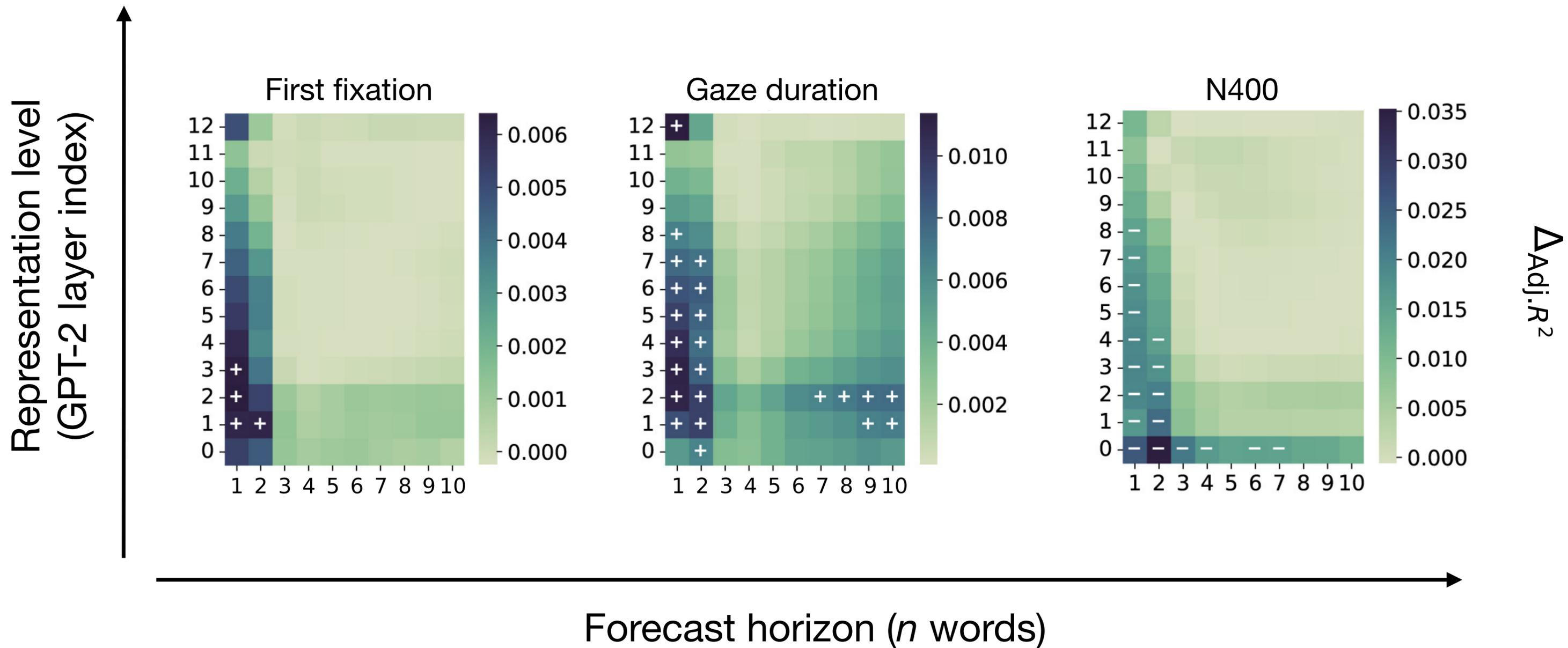


Giulianelli et al. (2026).
Incremental alternative sampling
as a lens into the temporal and
representational resolution of
linguistic prediction. *JML*.

Incremental alternative sampling

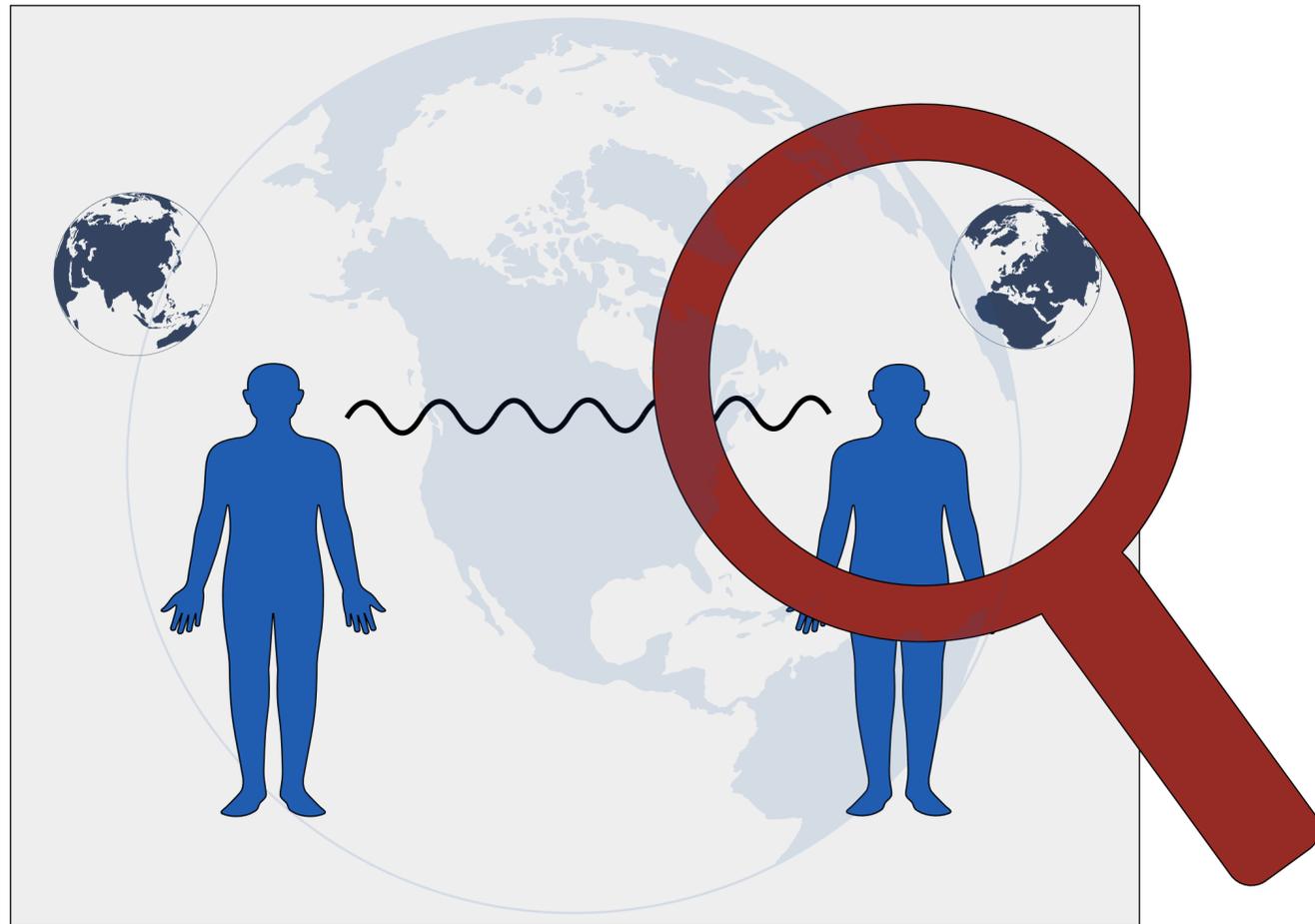


Incremental alternative sampling



If time permits

Language comprehension

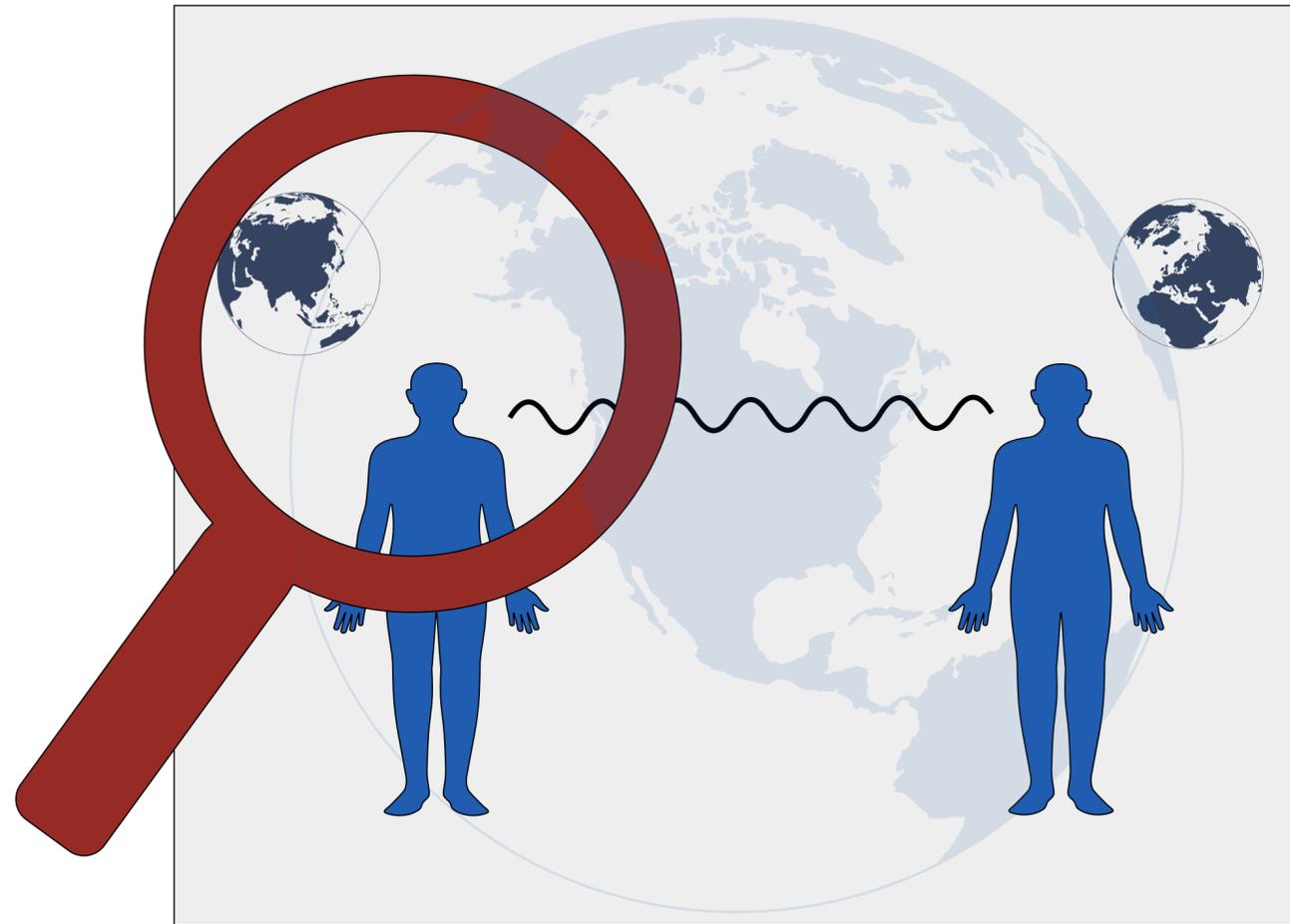


What is the role of prediction in language processing?

At which representational and temporal resolution does prediction take place?

Can behavioural and neural responses to language input be explained in terms of the input's information profile?

Language production

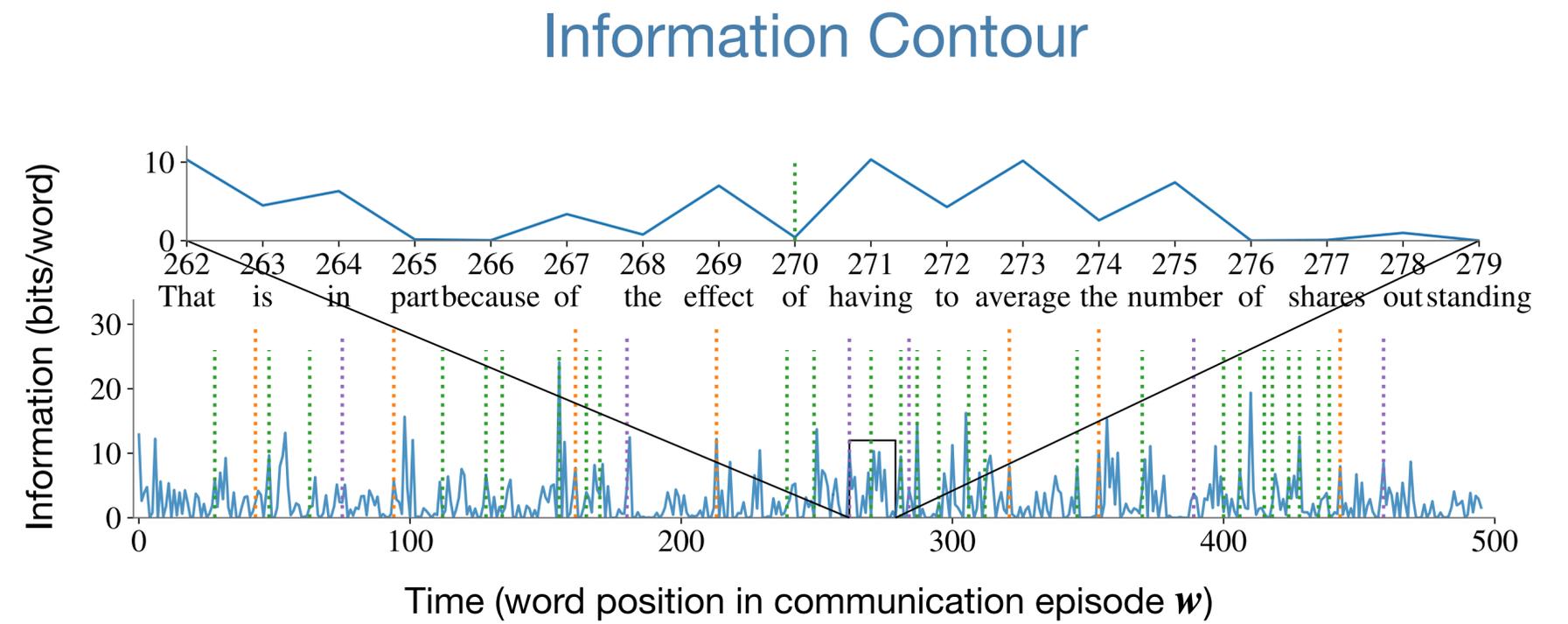
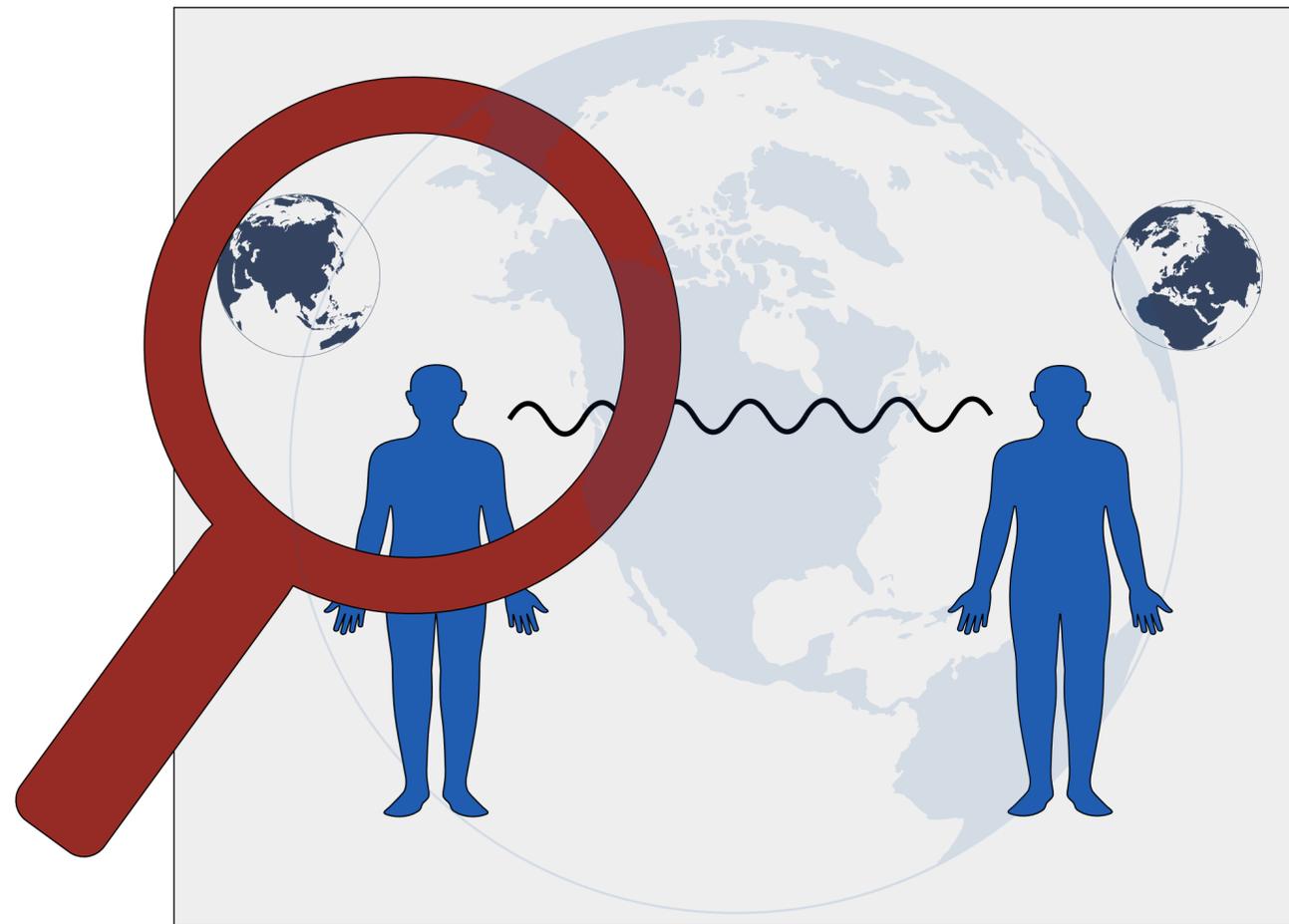


What is the rate at which producers transmit information?

Do producers make rational use of the communication channel?

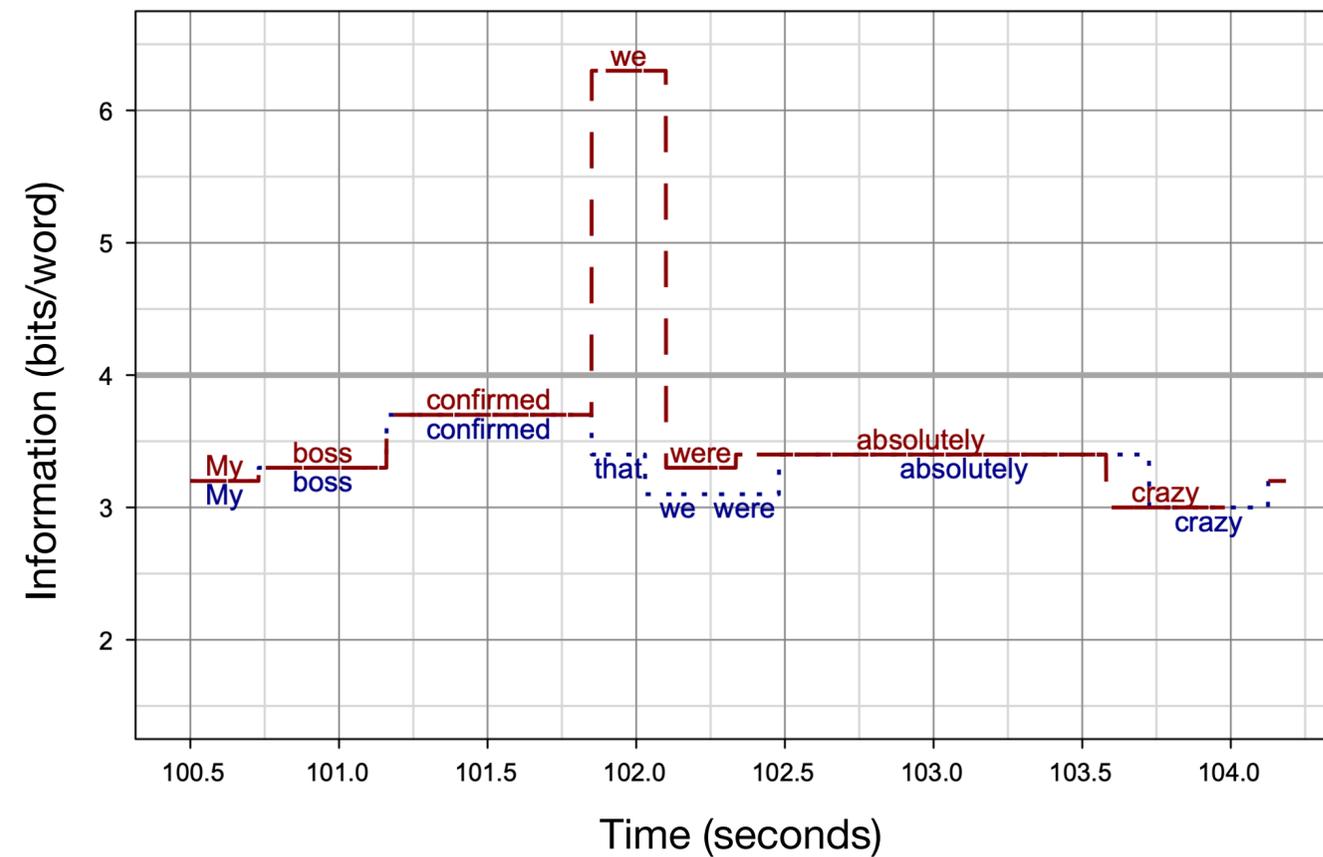
How does context (linguistic and non-linguistic) modulate information rate?

Information contours in texts and dialogues



$$-\log p(w_t | w_{<t})$$

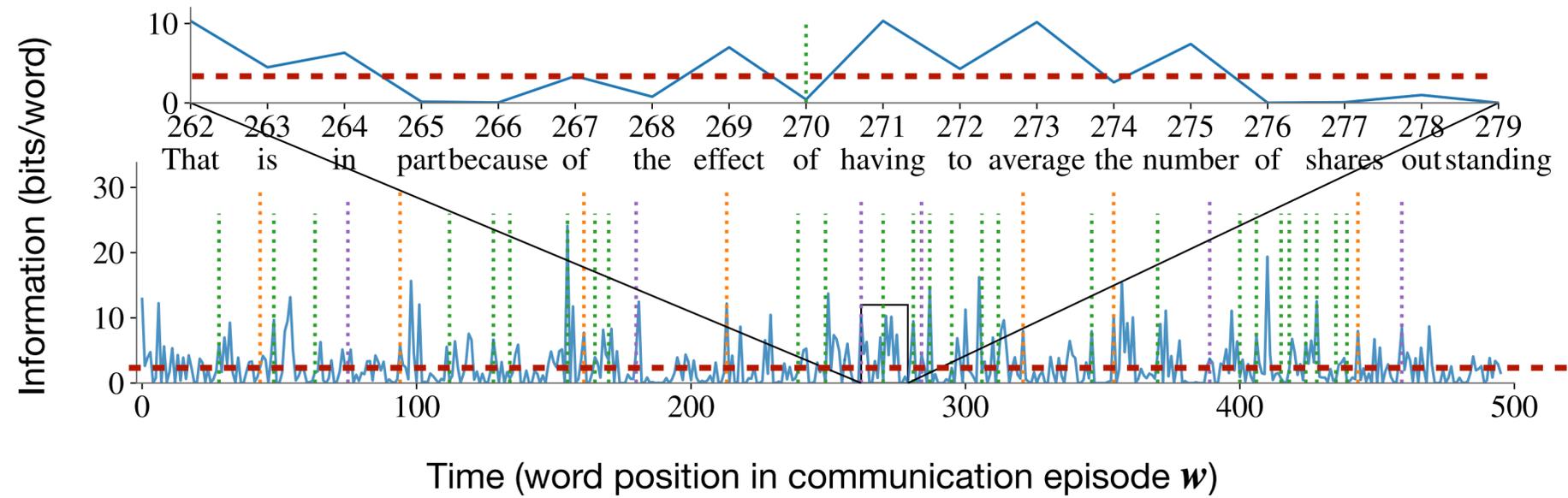
Information contours in texts and dialogues



Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

Information contours in texts and dialogues

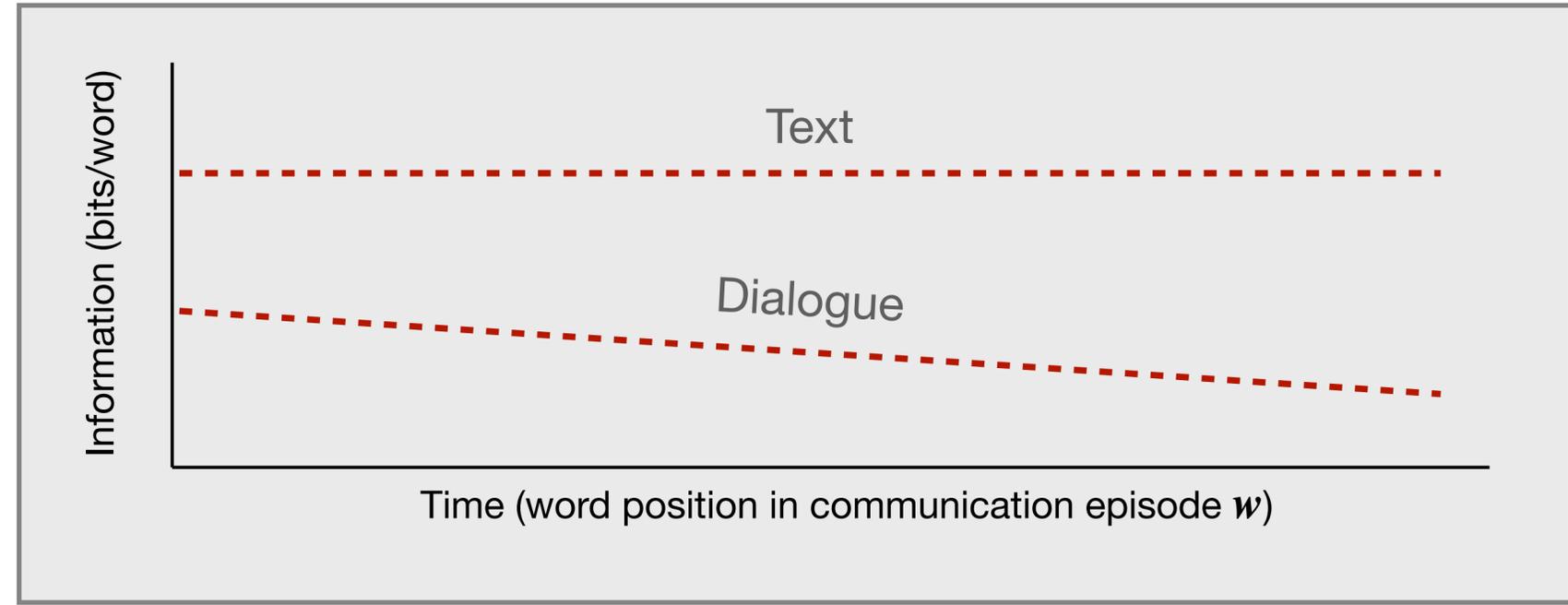


Hypothesis 1: Uniform Information Density

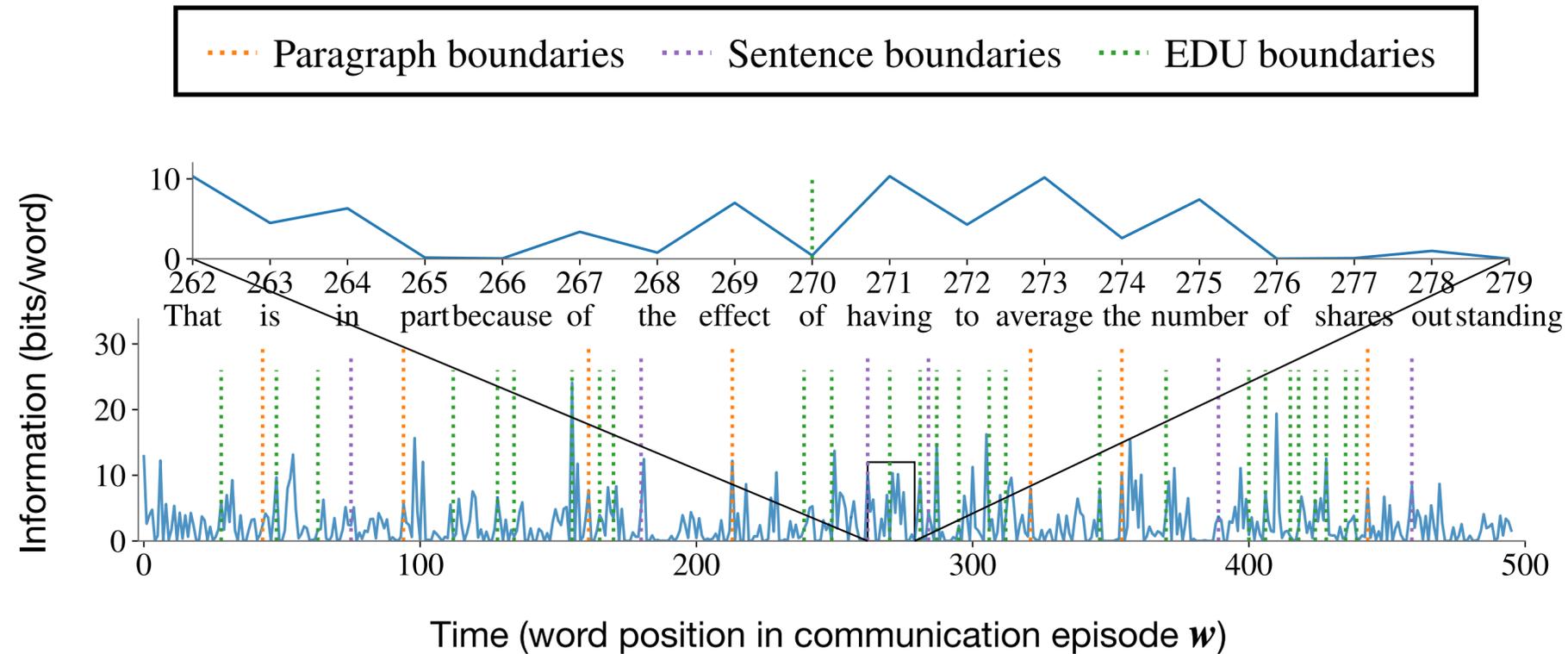
Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.



Information contours in texts and dialogues



Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.

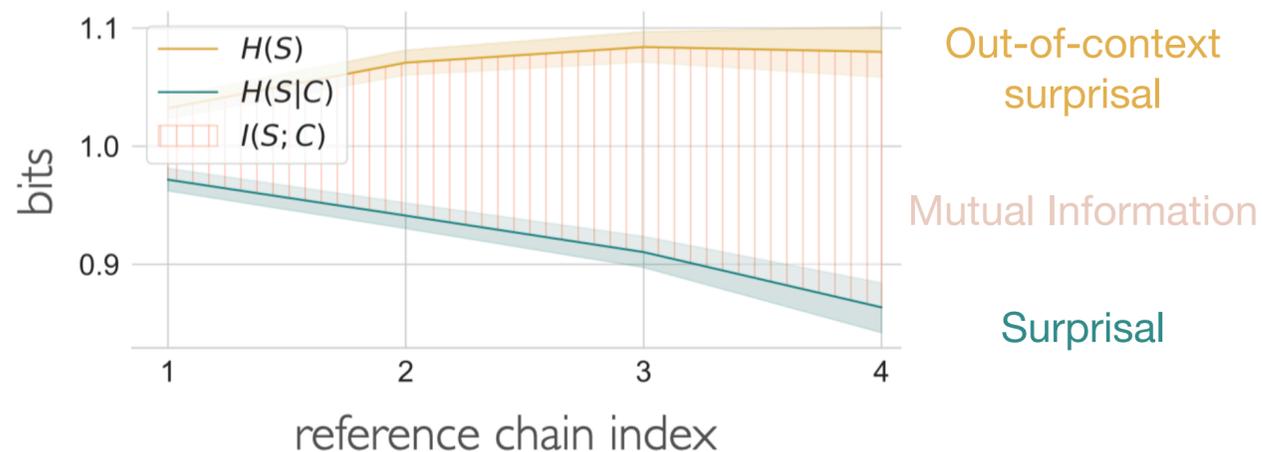
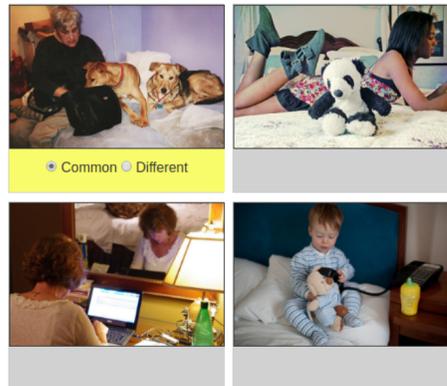
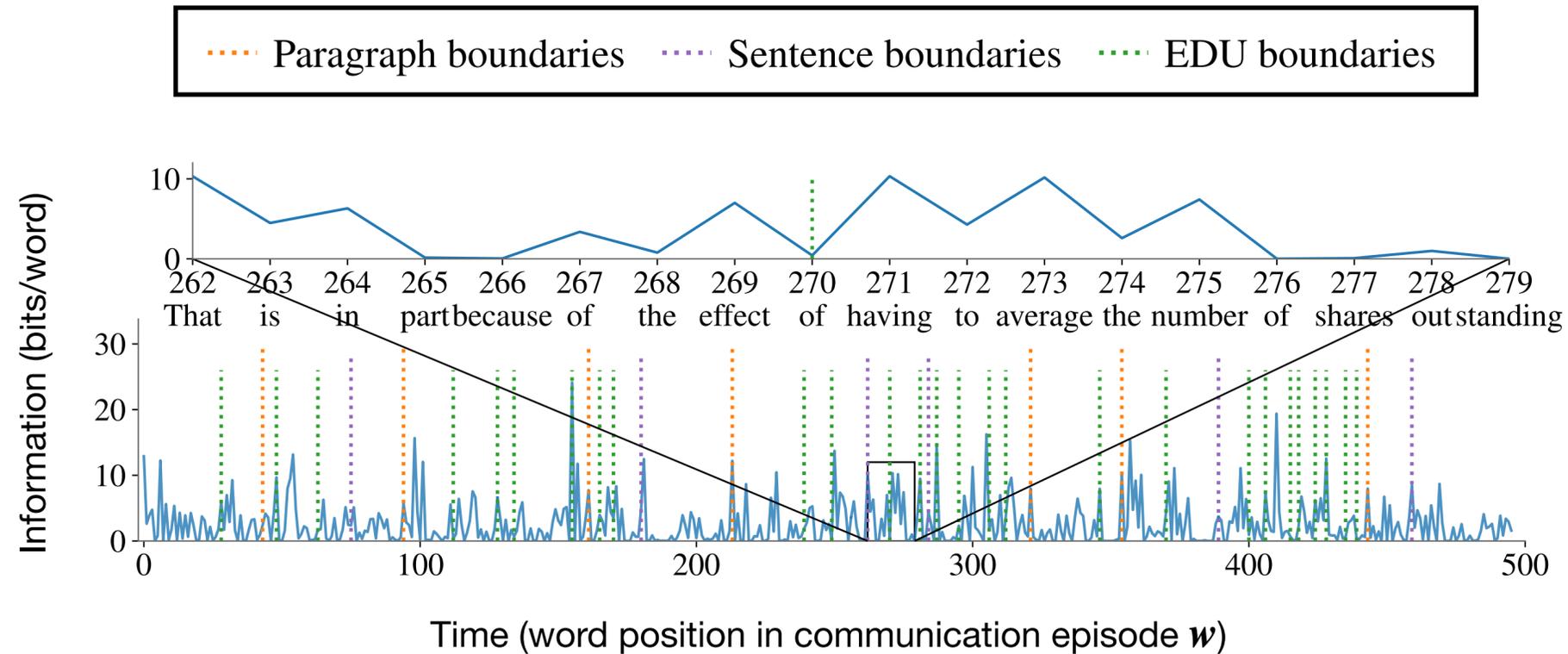
Hypothesis 2: Structured Context

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w are (partially) determined by the position of w_t within the hierarchy of w 's constituent structural units.

- a unit's position within **contextual structure** predicts its surprisal
 - RST discourse units in texts
 - task-specific contextual units in dialogues

Giulianelli, Sinclair, Fernández. EMNLP 2021.
Tsipidi, Nowak, Cotterell, Wilcox, Giulianelli, Warstadt. EMNLP 2024.

Information contours in texts and dialogues



Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.

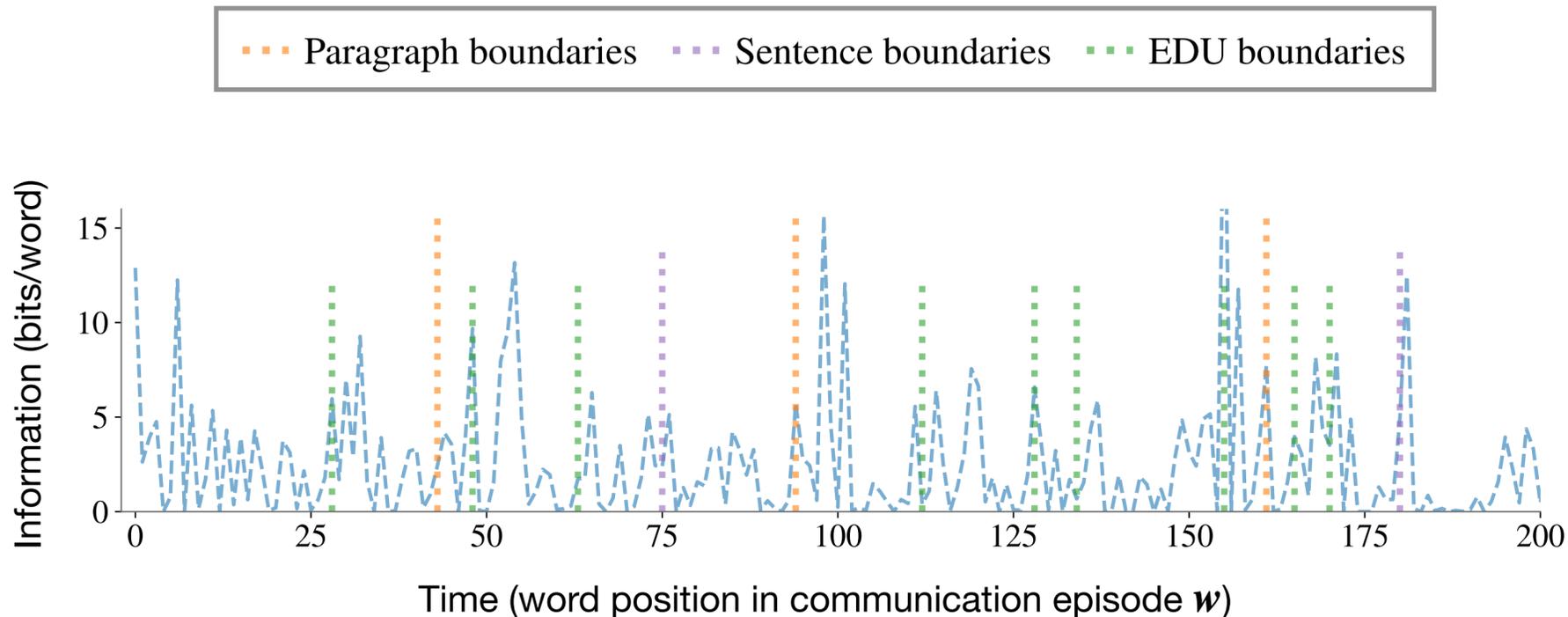
Hypothesis 2: Structured Context

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w are (partially) determined by the position of w_t within the hierarchy of w 's constituent structural units.

- a unit's position within **contextual structure** predicts its surprisal
- RST discourse units in texts
- task-specific contextual units in dialogues

Giulianelli, Sinclair, Fernández. EMNLP 2021.
Tsipidi, Nowak, Cotterell, Wilcox, Giulianelli, Warstadt. EMNLP 2024.

Information contours in texts and dialogues



Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.

Hypothesis 2: Structured Context

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w are (partially) determined by the position of w_t within the hierarchy of w 's constituent structural units.

- a unit's position within **contextual structure** predicts its surprisal
 - RST discourse units in texts
 - task-specific contextual units in dialogues

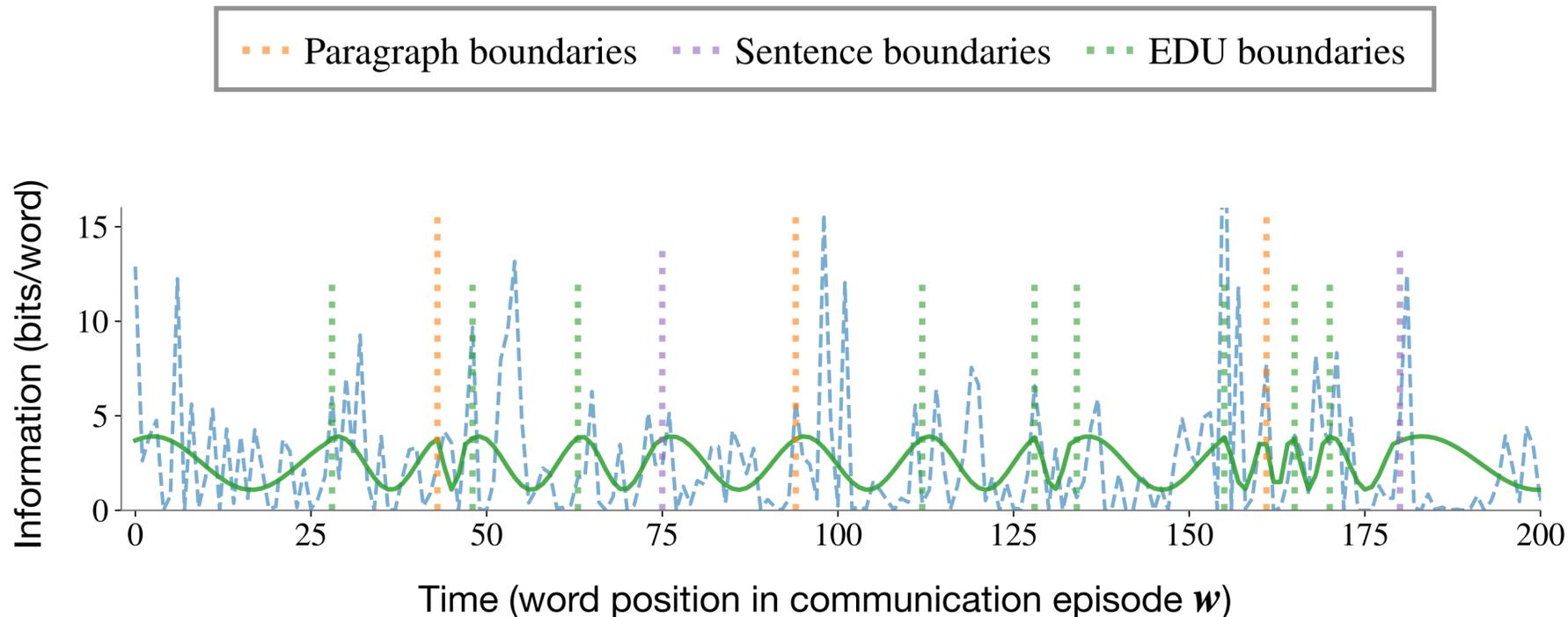
Giulianelli, Sinclair, Fernández. EMNLP 2021.
Tsipidi, Nowak, Cotterell, Wilcox, Giulianelli, Warstadt. EMNLP 2024.

Hypothesis 3: Harmonic Surprisal

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w vary periodically, with periods that correspond to the boundaries of structural units within w .

Tsipidi, Kiegeland, Nowak, Xu, Wilcox, Warstadt, Cotterell, Giulianelli.
ACL 2025.

Information contours in texts and dialogues



Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.

Hypothesis 2: Structured Context

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w are (partially) determined by the position of w_t within the hierarchy of w 's constituent structural units.

- a unit's position within **contextual structure** predicts its surprisal
 - RST discourse units in texts
 - task-specific contextual units in dialogues

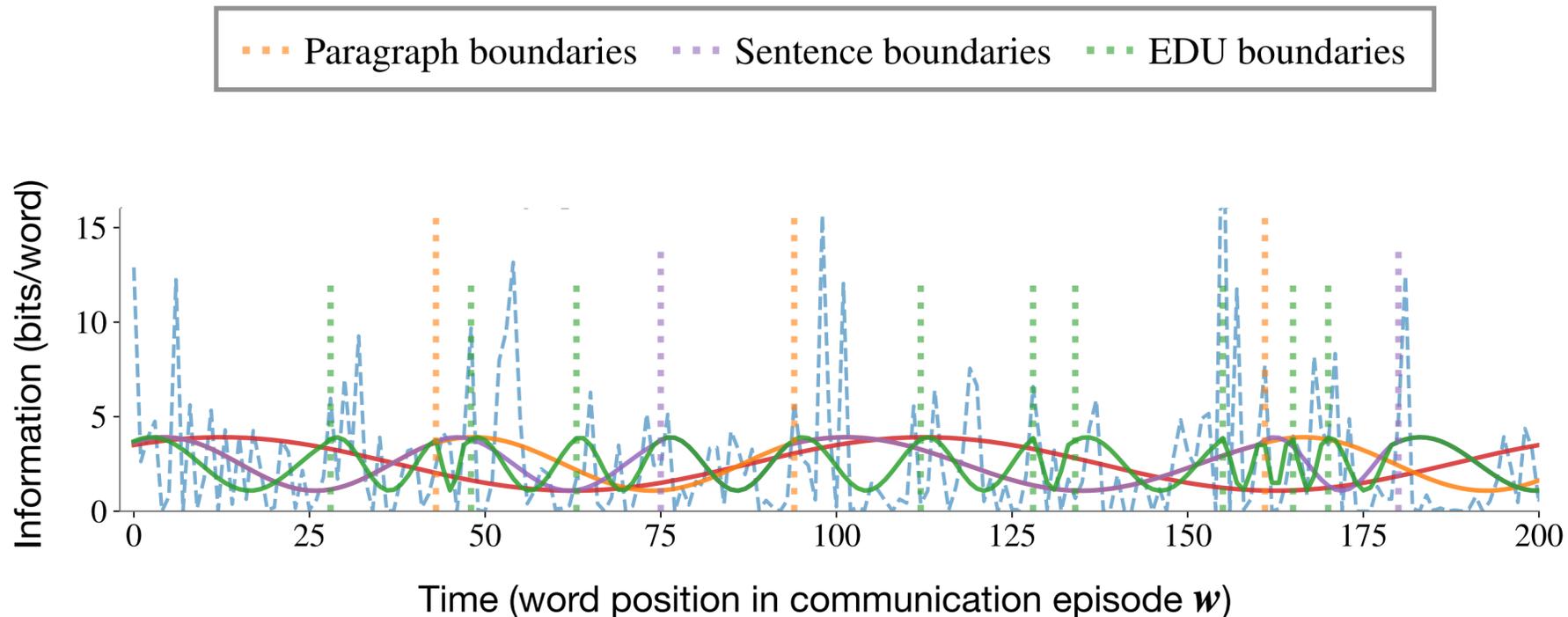
Giulianelli, Sinclair, Fernández. EMNLP 2021.
Tsipidi, Nowak, Cotterell, Wilcox, Giulianelli, Warstadt. EMNLP 2024.

Hypothesis 3: Harmonic Surprisal

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w vary periodically, with periods that correspond to the boundaries of structural units within w .

Tsipidi, Kiegeland, Nowak, Xu, Wilcox, Warstadt, Cotterell, Giulianelli.
ACL 2025.

Information contours in texts and dialogues



- - - Paragraph boundaries
 - - - Sentence boundaries
 - - - EDU boundaries

— Unscaled sinusoid — Paragraph-scaled sinusoid
— Sentence-scaled sinusoid — EDU-scaled sinusoid

Hypothesis 1: Uniform Information Density

Subject to the constraints of the grammar, speakers optimise their linguistic signals such that the surprisals l_w are distributed as uniformly as possible throughout a communication episode w .

- no evidence of local uniformity, pressure toward a global mean
- information rate decreases in dialogues

Giulianelli & Fernández. CoNLL 2021.
Giulianelli, Sinclair, Fernández. ACL 2021.

Hypothesis 2: Structured Context

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w are (partially) determined by the position of w_t within the hierarchy of w 's constituent structural units.

- a unit's position within **contextual structure** predicts its surprisal
- RST discourse units in texts
- task-specific contextual units in dialogues

Giulianelli, Sinclair, Fernández. EMNLP 2021.
Tsipidi, Nowak, Cotterell, Wilcox, Giulianelli, Warstadt. EMNLP 2024.

Hypothesis 3: Harmonic Surprisal

Values $l(w_t; w_{<t})$ in the surprisal contour l_w of a communication episode w vary periodically, with periods that correspond to the boundaries of structural units within w .

Tsipidi, Kiegeland, Nowak, Xu, Wilcox, Warstadt, Cotterell, Giulianelli.
ACL 2025.

Information contours in texts and dialogues

Producers' communicative strategies through the lens of information rate modulation

- facilitating production (e.g., repetitions)

Giulianelli, Sinclair, Fernández. ACL 2022.

- enhancing coordination in dialogue

Yee, Giulianelli, Sinclair. LREC-COLING 2024.

- style, genre, and writing quality

- human-generated vs. model-generated texts

- facilitating comprehension

- in multimodal contexts

Gay, Haley, Giulianelli, Ponto. EACL 2026.

