EXPECTATION-BASED SEMANTICS IN LANGUAGE COMPREHENSION

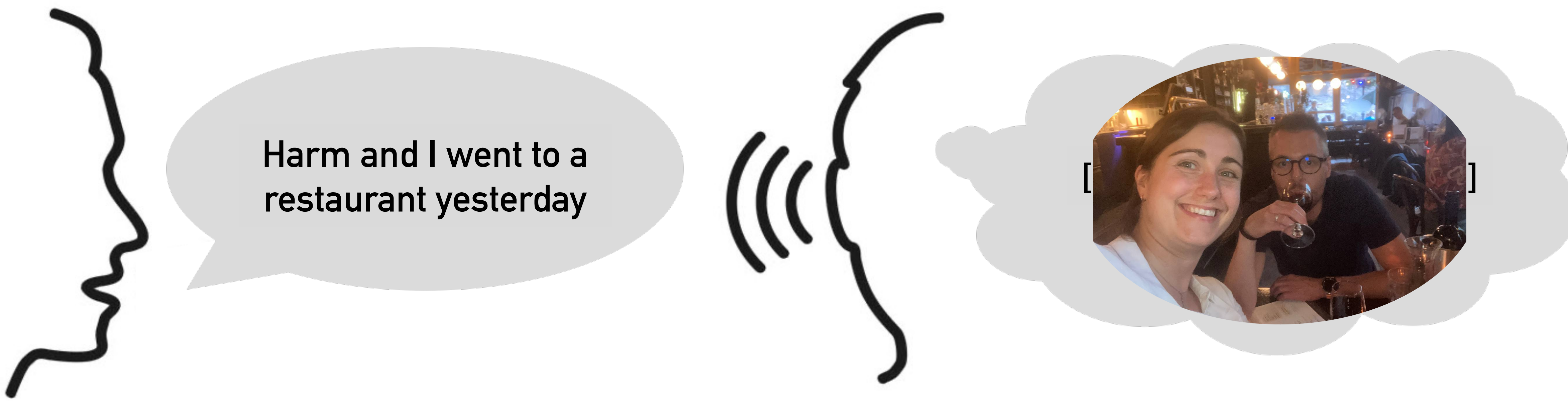
Noortje Venhuizen

Joint work with: Harm Brouwer, Matthew W. Crocker and Petra Hendriks

UNIVERSITÄT DES SAARLANDES

TILBURG UNIVERSITY

SEMANTICS IN HUMAN LANGUAGE COMPREHENSION



Big Q: What does this mental meaning representation look like?

Representation of mental meaning should be:

- *expressive and compositional*: compose logically complex meanings
- *graded and inferential*: probabilistic inferences beyond literal meaning
- *incremental*: word-by-word construction of meaning
- *neurally plausible*: allow for implementation in neural hardware

Brouwer, Crocker, Venhuizen (2017, From Semantics to Dialectometry)

FROM FORMAL MODELS TO MEANING SPACE

$M_1 = \langle U_1, V_1 \rangle$

$p_1 \land \neg p_2 \land p_3 \land \ldots$

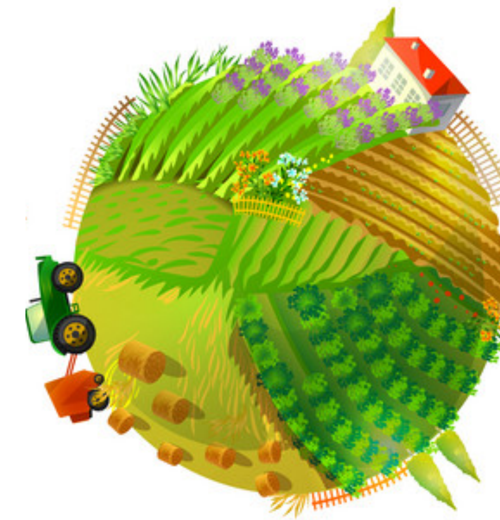
$M_2 = \langle U_2, V_2 \rangle$

$\neg p_1 \land p_2 \land p_3 \land \ldots$

$M_3 = \langle U_3, V_3 \rangle$

$\neg p_1 \land p_2 \land \neg p_3 \land \ldots$

…



$M_n = \langle U_n, V_n \rangle$

$\neg p_1 \land \neg p_2 \land \neg p_3 \land \ldots$

- Individual models describe states-of-affairs over all propositions in $\mathcal{P}$
- The set of models $\mathcal{M}_{\mathcal{P}}$ defines a meaning space
- Propositional meaning defined by co-occurrence across models (~ world knowledge)

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2022, Inf. Comp.)

DISTRIBUTIONAL FORMAL SEMANTICS

propositional meaning vectors

formal models

	p_1	p_2	p_3	p_4	…
M_1	1	1	0	0	…
M_2	1	0	0	1	…
M_3	0	1	0	1	…
M_4	1	1	1	1	…
M_5	0	1	0	0	…
…	…	…	…	…	…

$$[\![p_j]\!]^{\mathcal{M}} := v(p_j)$$

$$\textit{where: } v_i(p_j) = 1 \textit{ iff } M_i \vDash p_j$$

- **Incremental inference-based probabilistic sampling:** Based on a set of propositions $\mathcal{P}$, we sample a set of models $\mathcal{M}_{\mathcal{P}}$—taking into account hard and probabilistic world knowledge constraints
- **Co-occurrence defines meaning:** Propositions with related meanings are true in many of the same models, resulting in similar meaning vectors

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2022, Inf. Comp.)

THE DISTRIBUTIONAL HYPOTHESIS REVISITED

> You shall know a word
> by the company it keeps

- J. R. Firth (1957)

THE DISTRIBUTIONAL HYPOTHESIS REVISITED

"

You shall know a ~~word~~ *proposition*
by the company it keeps *(in the world)*

- J. R. Firth (1957)

MEANING VECTOR COMPOSITION

Meaning vectors can be combined to define compositional meanings

- Standard logical operators interpreted as in model-theory

$$v_i(\neg p) = 1 \quad \textit{iff} \quad M_i \nvDash p$$

$$v_i(p \wedge q) = 1 \quad \textit{iff} \quad M_i \vDash p \textit{ and } M_i \vDash q$$

… etc.

- Quantification is defined relative to the combined universe of $\mathcal{M}_\mathcal{P}$: $\mathcal{U}_\mathcal{M} = \{e_{1 \ldots} e_m\}$ (thereby preserving entailment in $\mathcal{M}_\mathcal{P}$)

$$v_i(\forall x \varphi) = 1 \quad \textit{iff} \quad M_i \vDash \varphi[x \backslash e_1] \wedge \ldots \wedge \varphi[x \backslash e_m]$$

$$v_i(\exists x \varphi) = 1 \quad \textit{iff} \quad M_i \vDash \varphi[x \backslash e_1] \vee \ldots \vee \varphi[x \backslash e_m]$$

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2022, Inf. Comp.)

PROBABILITIES IN THE MEANING SPACE

All (sub-)propositional meaning vectors inherently encode (co-)occurrence probabilities

- Prior probability of meaning vector *a*

$$P(a) = \frac{1}{|\mathcal{M}|}\sum_i \vec{v}_i(a)$$

- Conjunction probability between *a* and *b*

$$P(a \wedge b) = \frac{1}{|\mathcal{M}|}\sum_i \vec{v}_i(a)\vec{v}_i(b)$$

- Conditional probability of *a* given *b*

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

	p_1	p_2	p_3	p_4	
M_1	1	1	0	0	...
M_2	1	0	0	1	...
M_3	0	1	0	1	...
M_4	1	1	1	1	...
	0	1	0	0	...
	...	...	...	...	...

QUANTIFYING PROBABILISTIC INFERENCE

Probabilistic logical inference of meaning vector *a* given *b*

$$inference(a,b) = \begin{cases} [P(a|b) - P(a)] \,/\, [1 - P(a)] & \text{if } P(a|b) > P(a) \\ [P(a|b) - P(a)] \,/\, P(a) & \text{otherwise} \end{cases}$$

- $P(a|b) > P(a)$: Positive inference (b increases probability of a)

$$inference(a,b) = 1 \Leftrightarrow b \vDash a$$

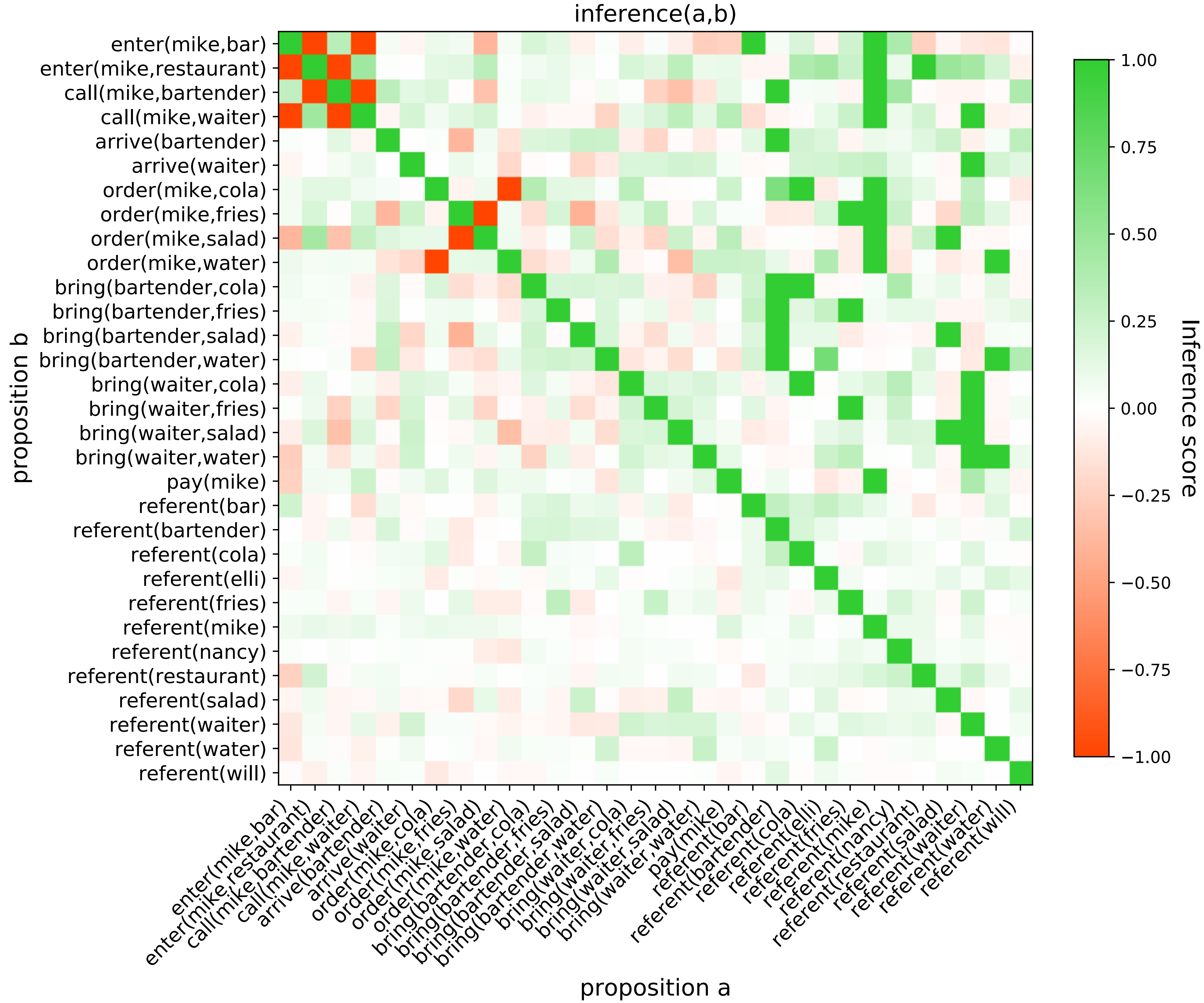
- $P(a|b) \leq P(a)$: Negative inference (b decreases probability of a)

$$inference(a,b) = -1 \Leftrightarrow b \vDash \neg a$$

Frank et al. (2009,Cognition)

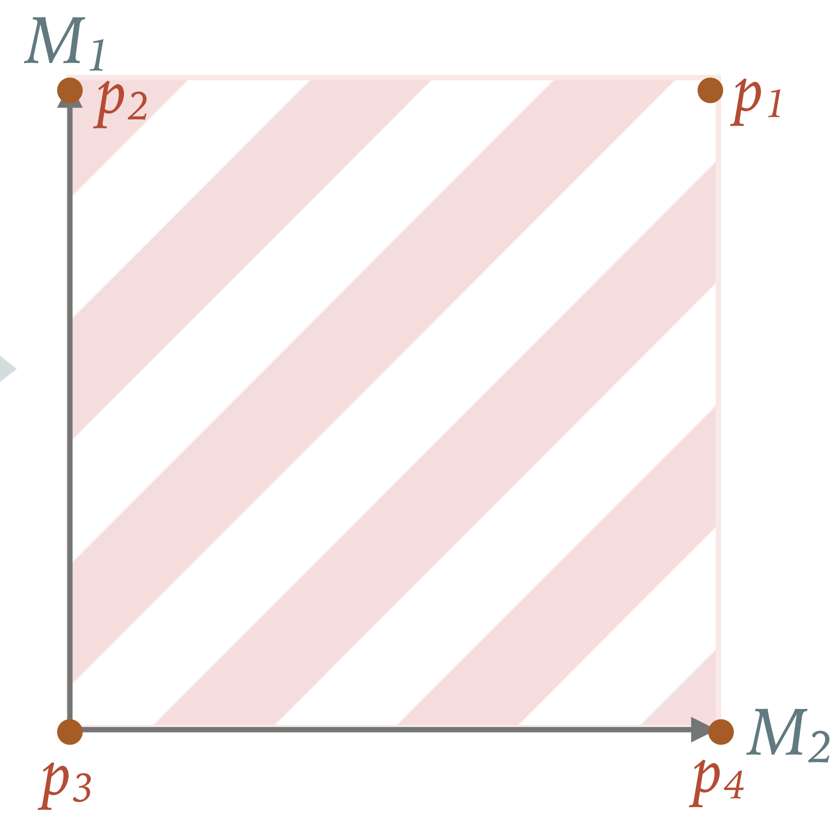
WORLD KNOWLEDGE IN THE MEANING SPACE

We sampled a meaning space of 150 models describing 51 propositions



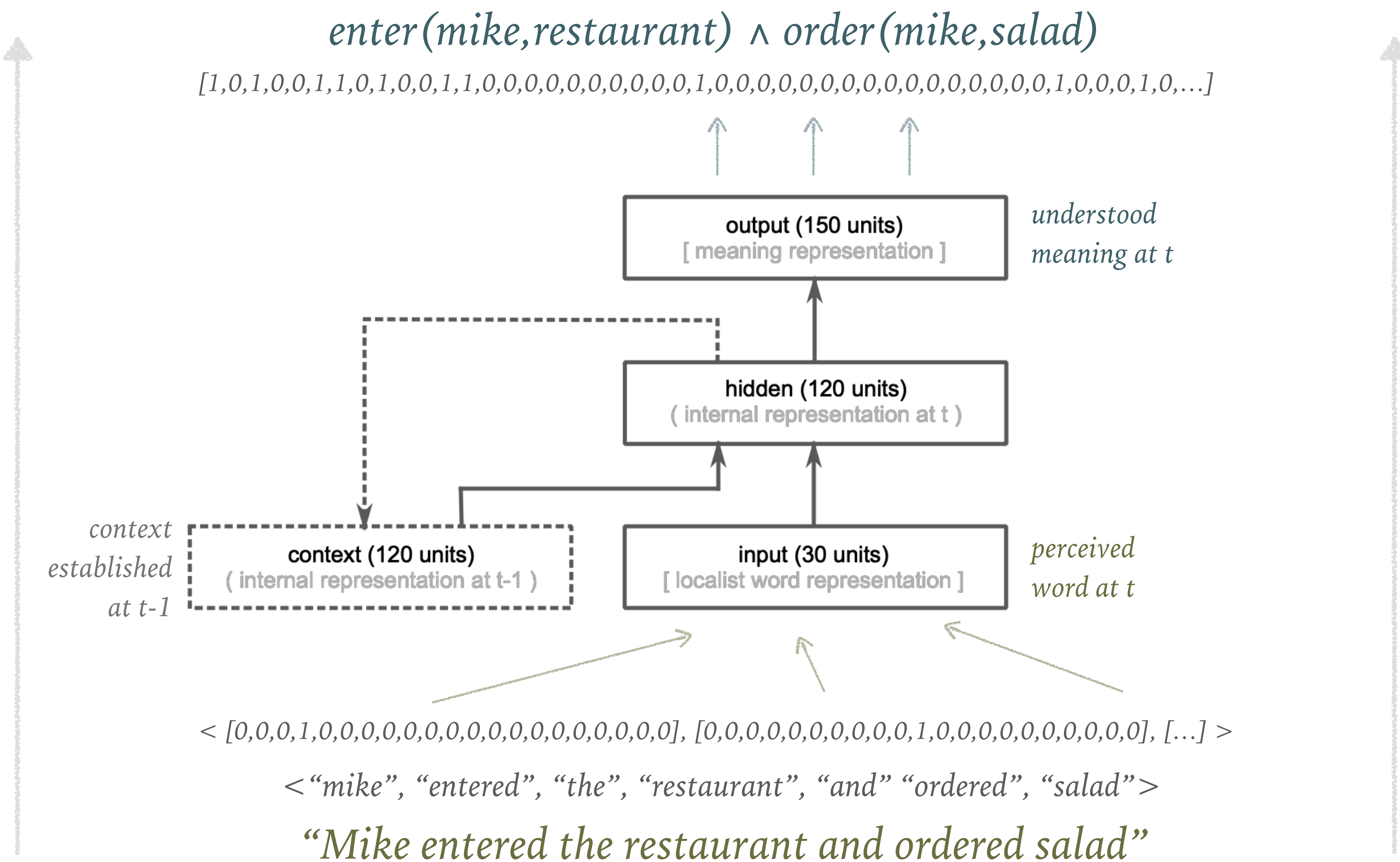
CONTINUOUS NATURE OF THE MEANING SPACE

	p_1	p_2	p_3	p_4	
M_1	1	1	0	0	…
M_2	1	0	0	1	…
M_3	0	1	0	1	…
M_4	1	1	1	1	…
M_5	0	1	0	0	…
	…	…	…	…	…



- Each point in the meaning space can be interpreted relative to $\mathcal{M}_{\mathcal{P}}$
  - Binary vectors: propositional meanings (simple or complex)
  - Real-valued vectors: sub-propositional meanings
- Sub-propositional meaning derives from incremental mapping from (sequences of) words to proposition-level meanings

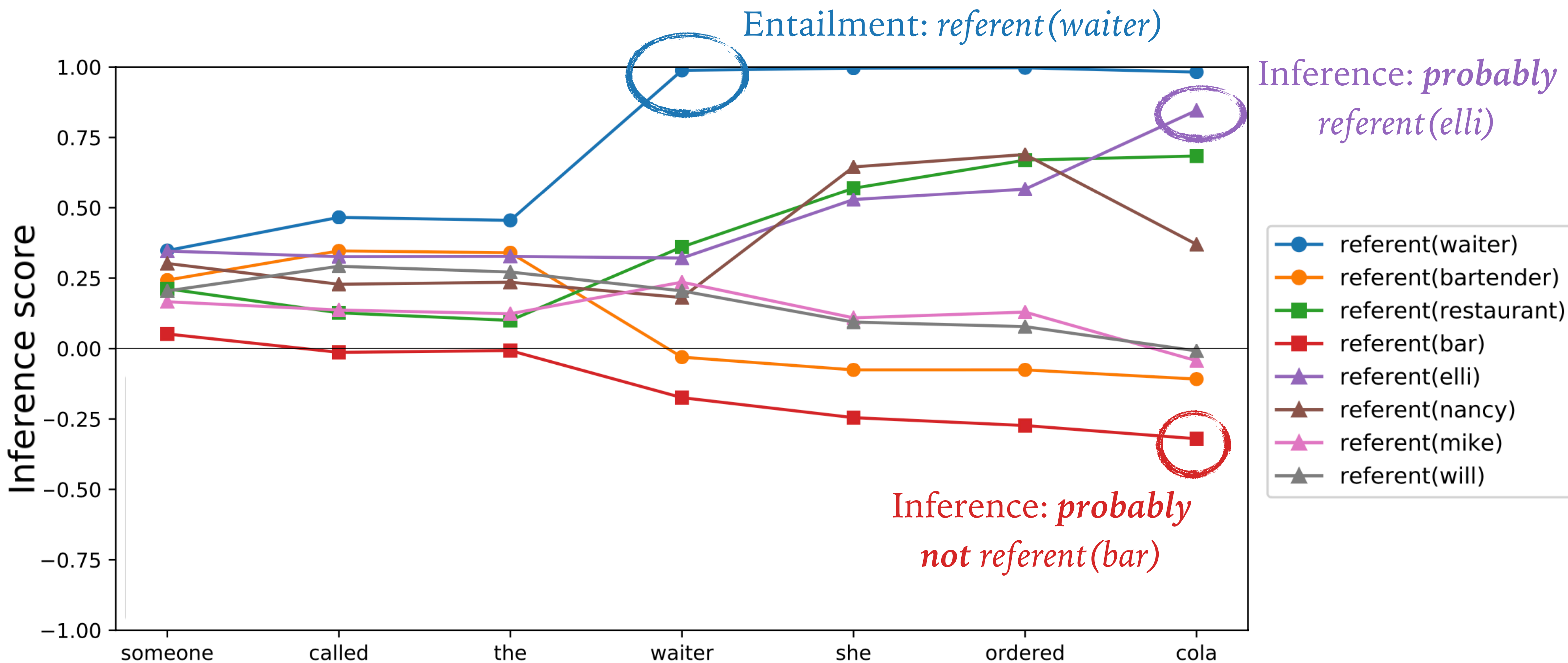
A MODEL OF INCREMENTAL MEANING CONSTRUCTION



ENTAILMENT AND INFERENCE

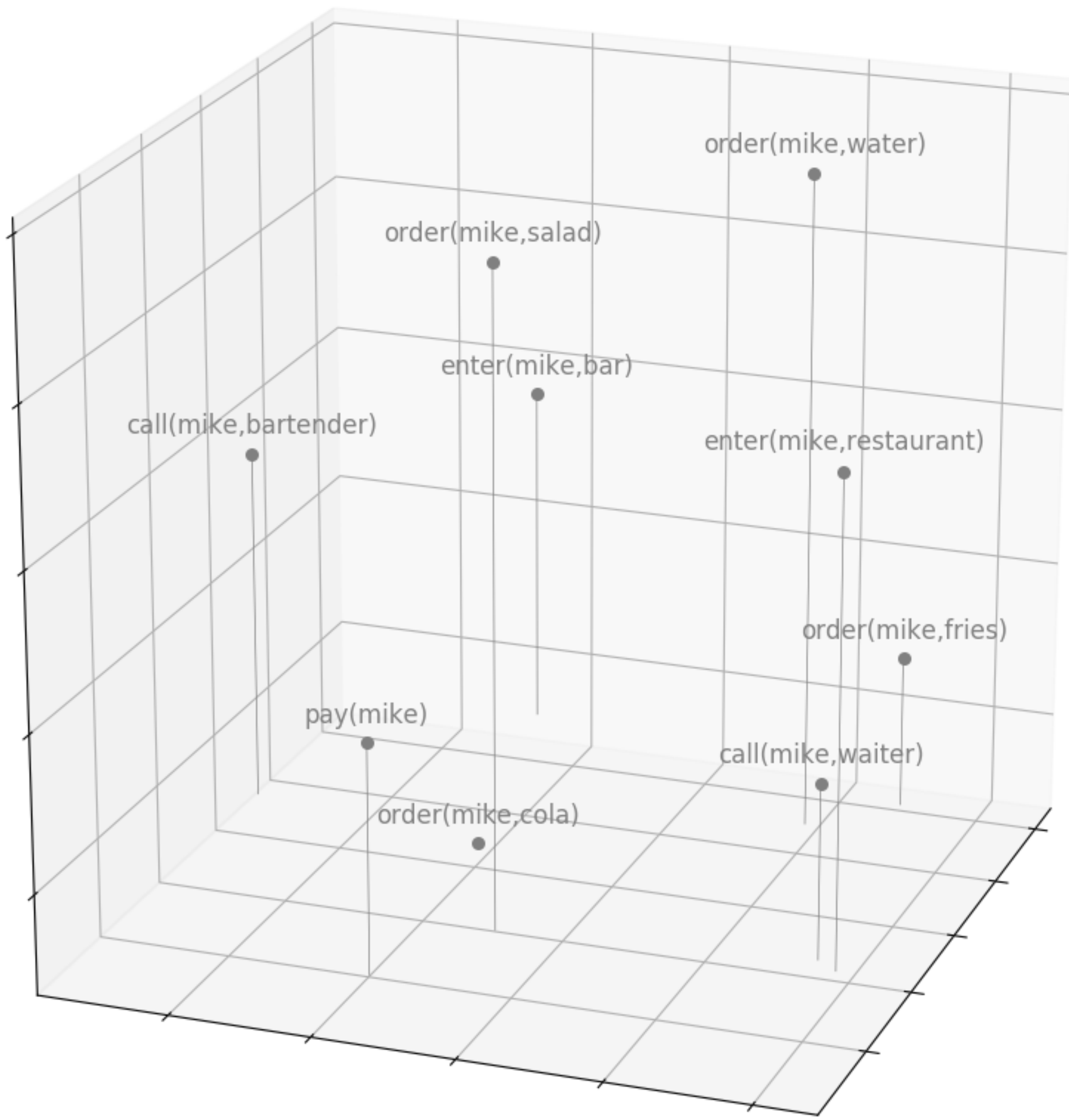
Incremental meaning construction in the model is driven by:

- ➤ Sentence-semantics mappings (linguistic experience)
- ➤ Structure of the meaning space (world knowledge)



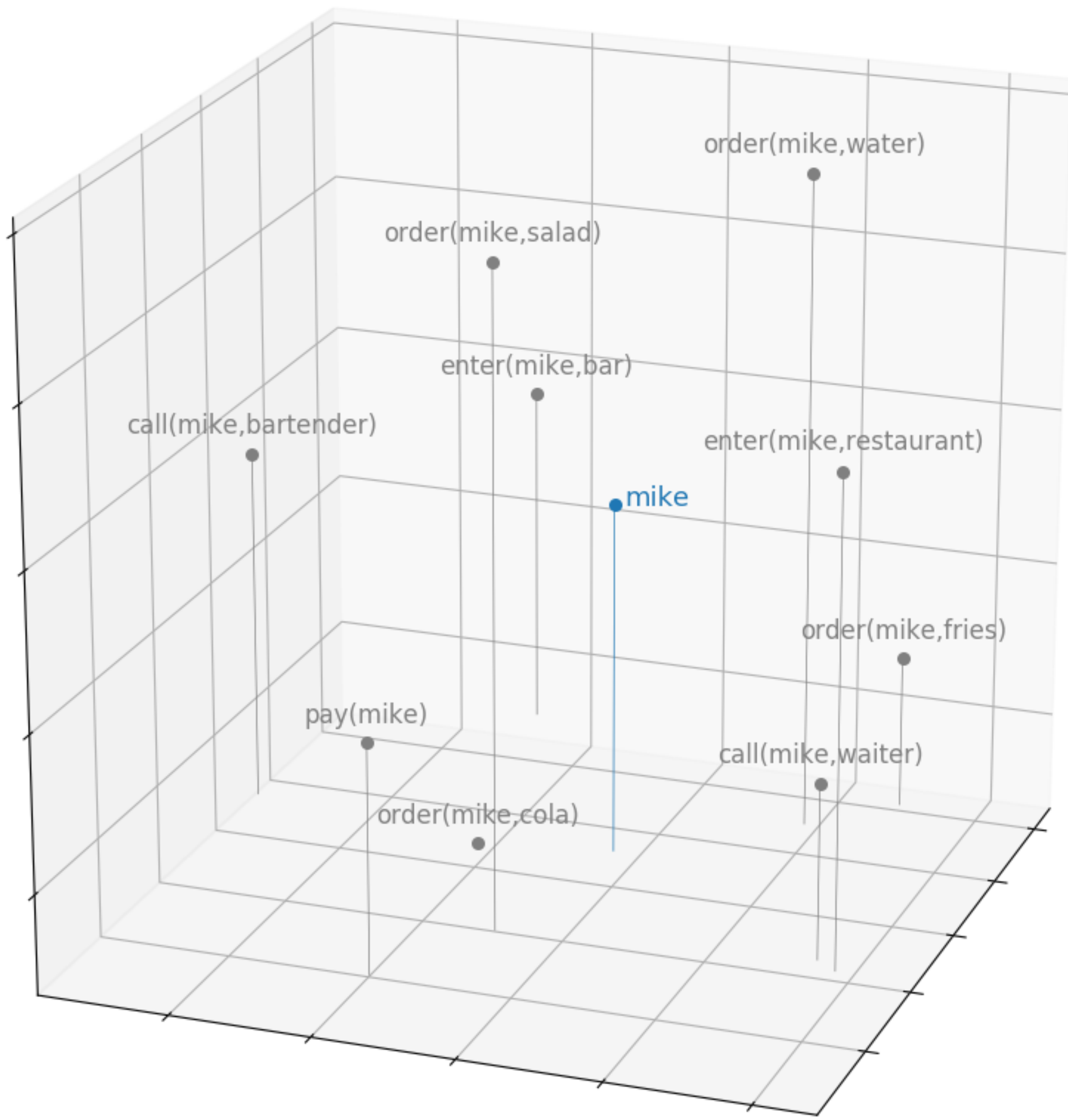
Venhuizen et al. (2022, Inf. Comp.)

MEANING SPACE NAVIGATION



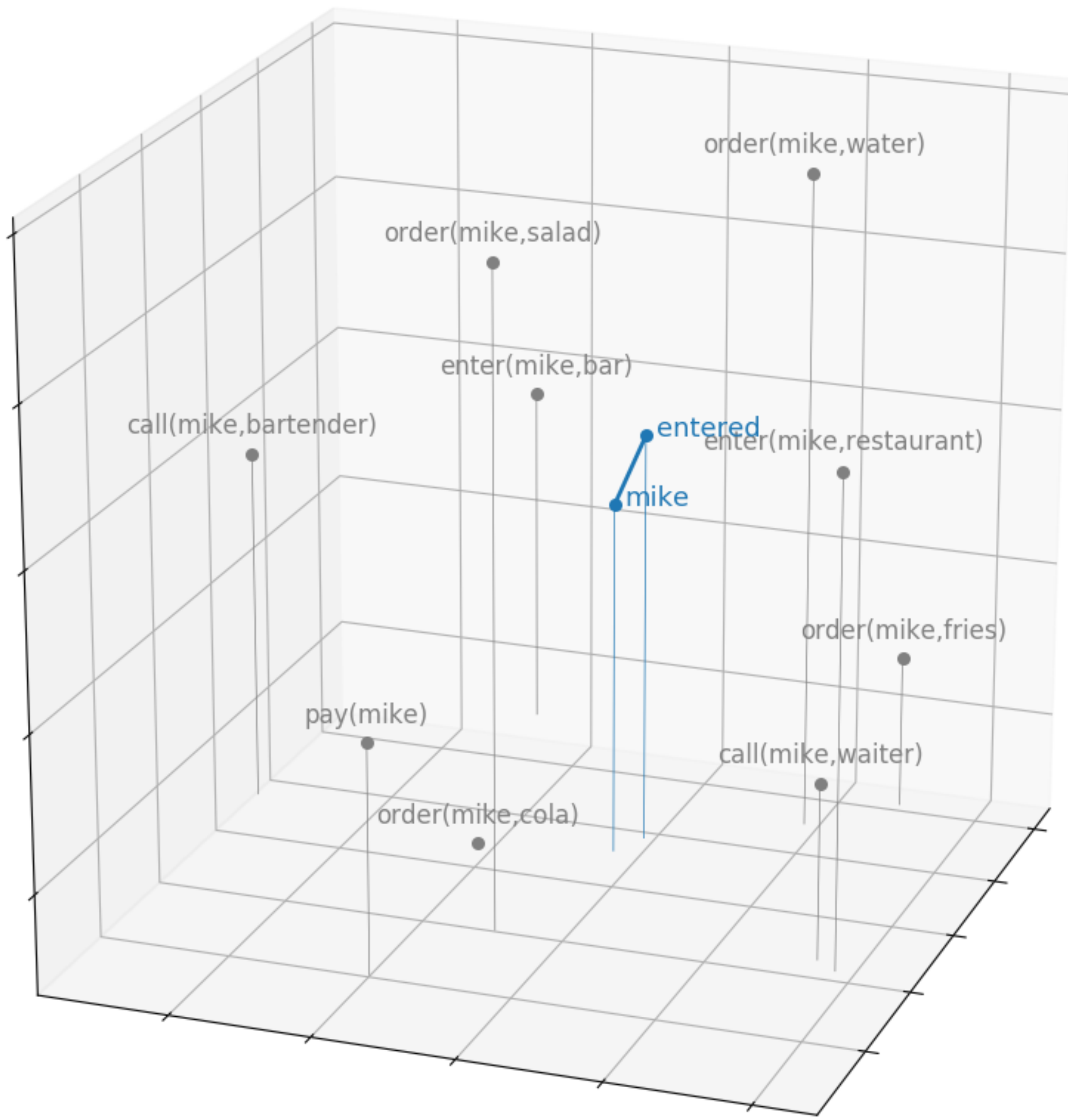
➤ 3-Dimensional representation of 150-D meaning space (MDS)

MEANING SPACE NAVIGATION



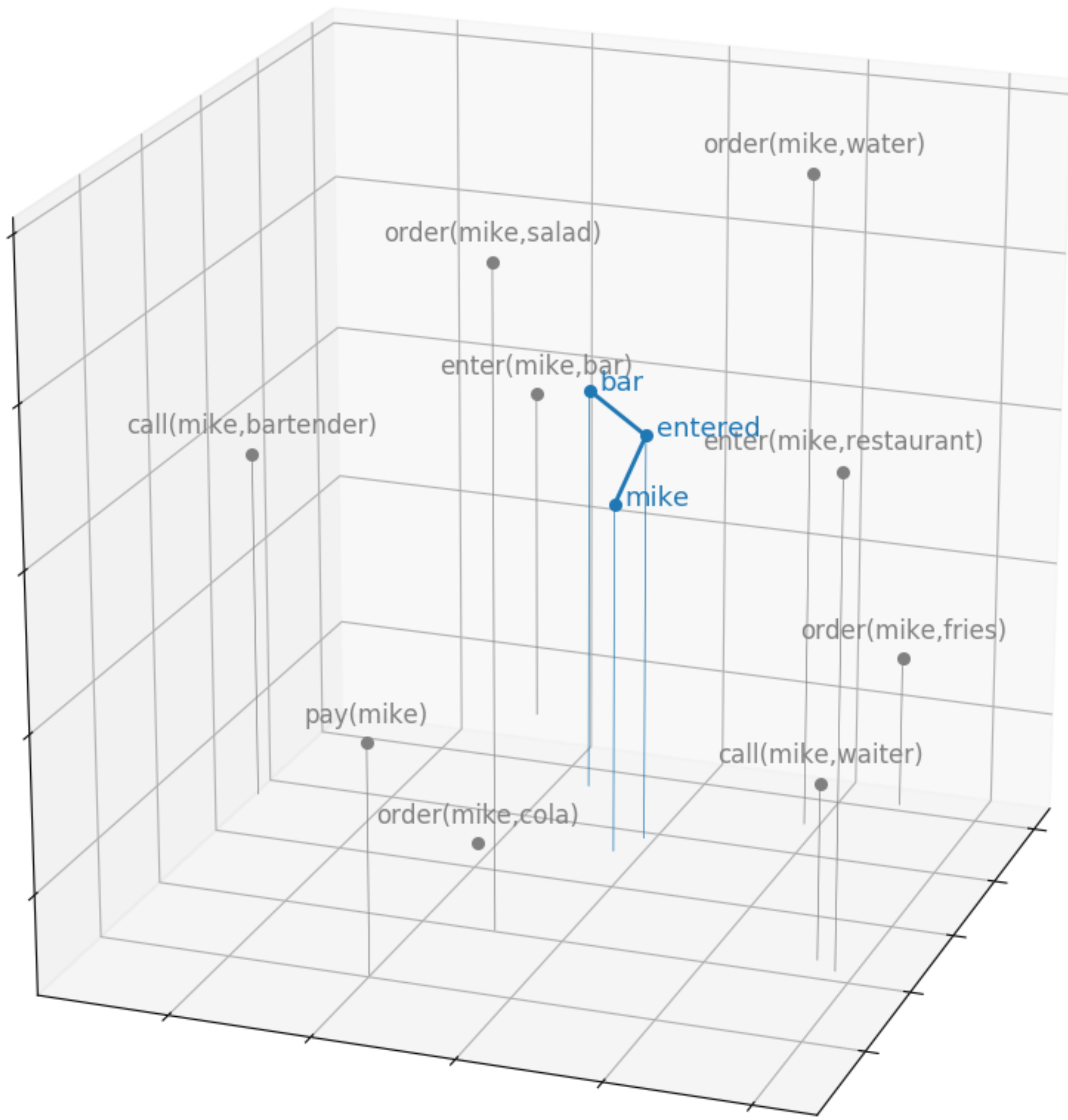
➤ Model-derived meaning of “Mike"

MEANING SPACE NAVIGATION



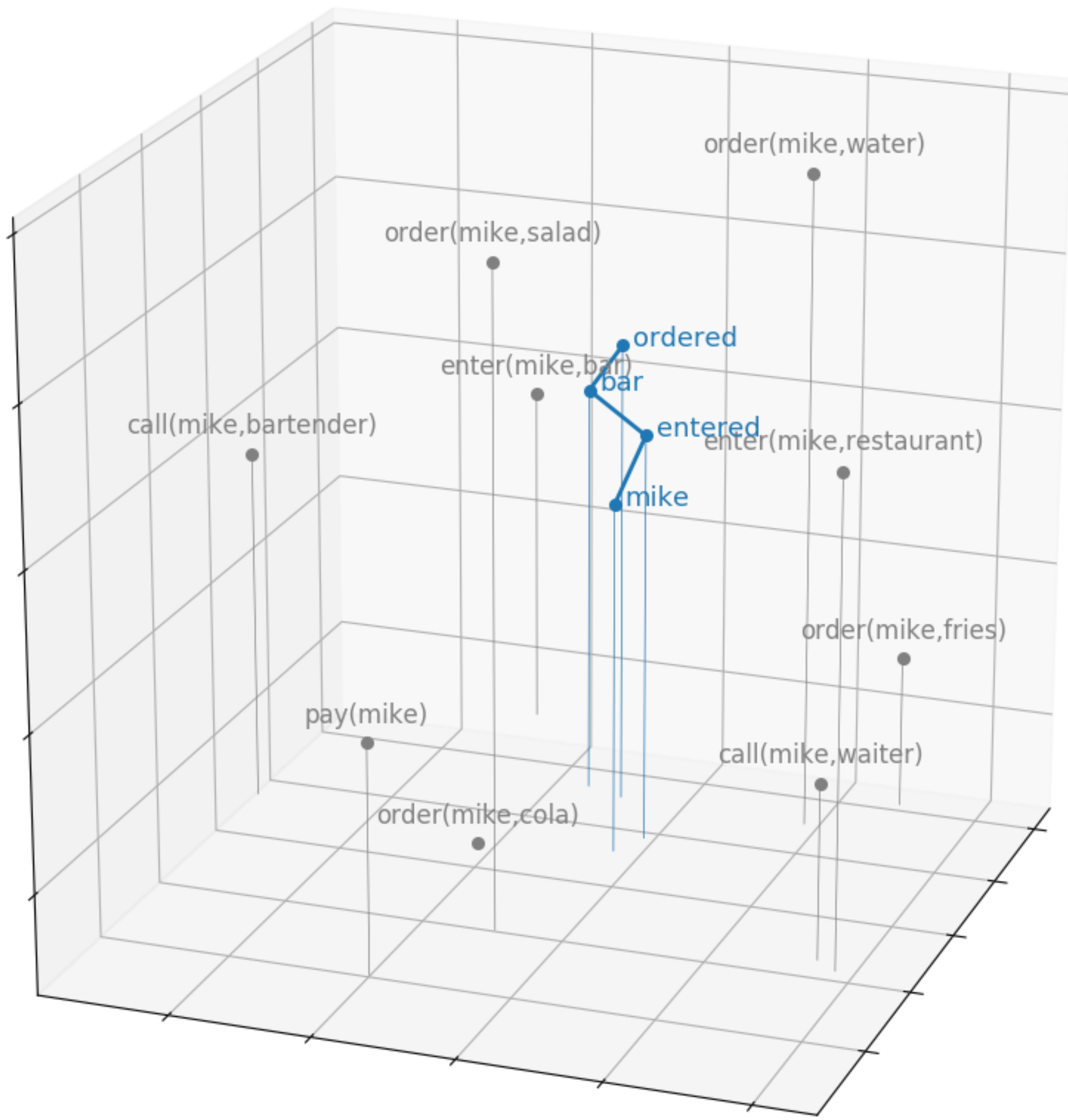
- Model-derived meaning of “Mike entered”

MEANING SPACE NAVIGATION



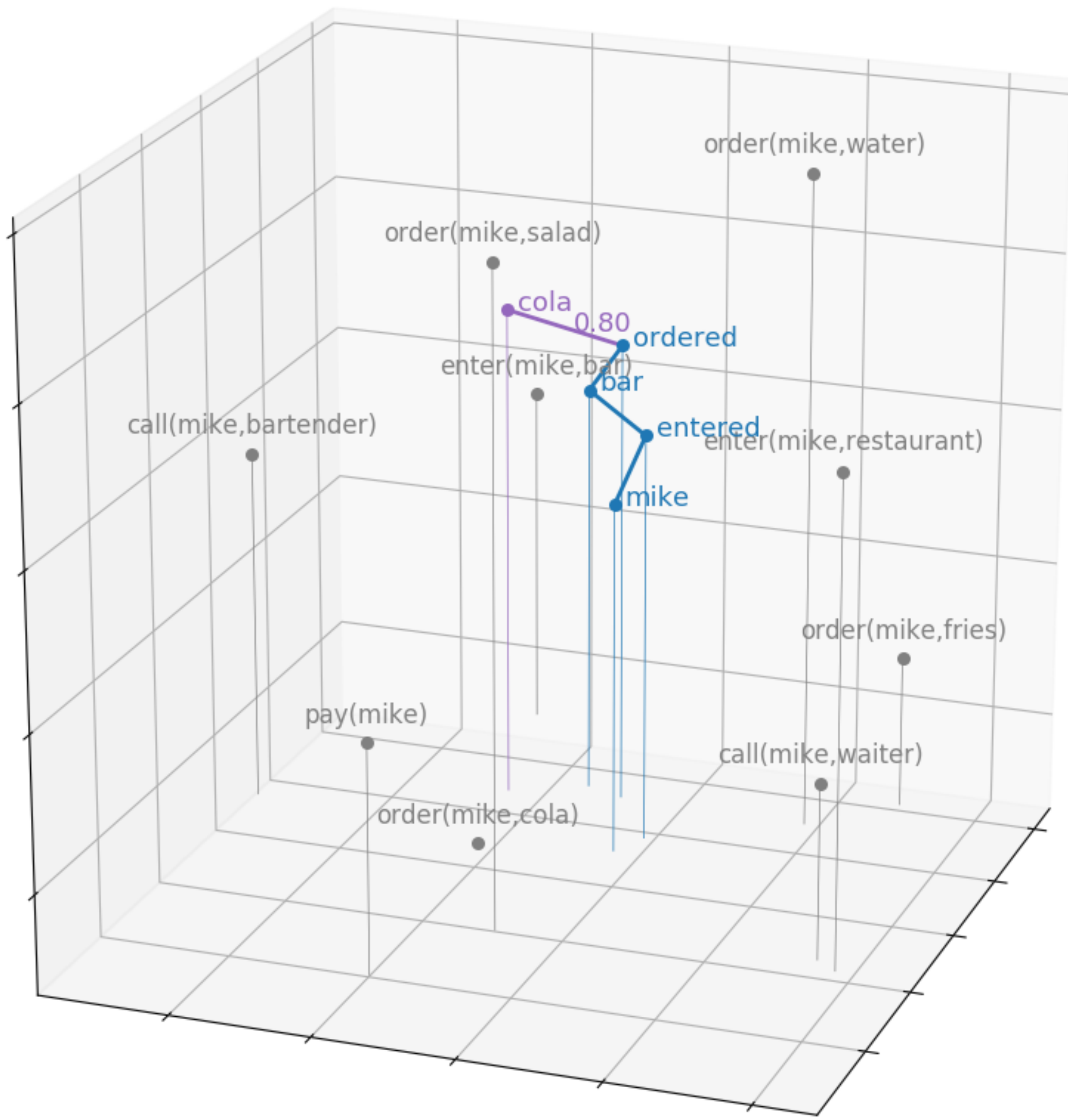
- Model-derived meaning of “Mike entered the bar”

MEANING SPACE NAVIGATION



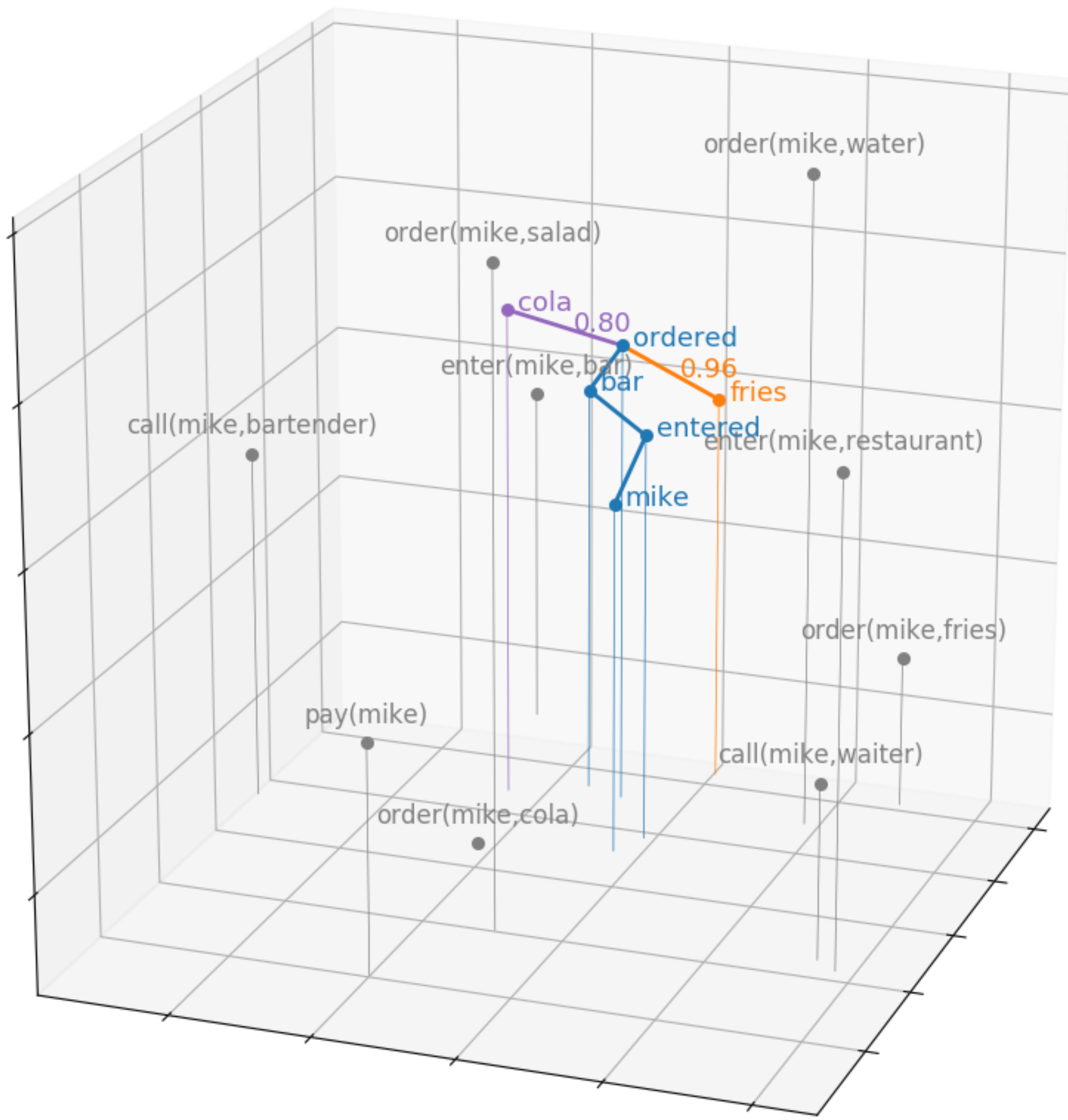
➤ Model-derived meaning of “Mike entered the bar [.] he ordered”

MEANING SPACE NAVIGATION



- Model-derived meaning of “Mike entered the bar [.] he ordered cola”
- Transition in meaning space quantifies expectancy of the continuation

MEANING SPACE NAVIGATION



- Model-derived meaning of “Mike entered the bar [.] he ordered fries”
- Larger (more surprising) transition reflects less expected continuation

INFORMATION THEORY IN DFS

Probabilistic nature of meaning space allows for defining formal notion of *information* (Shannon, 1948)

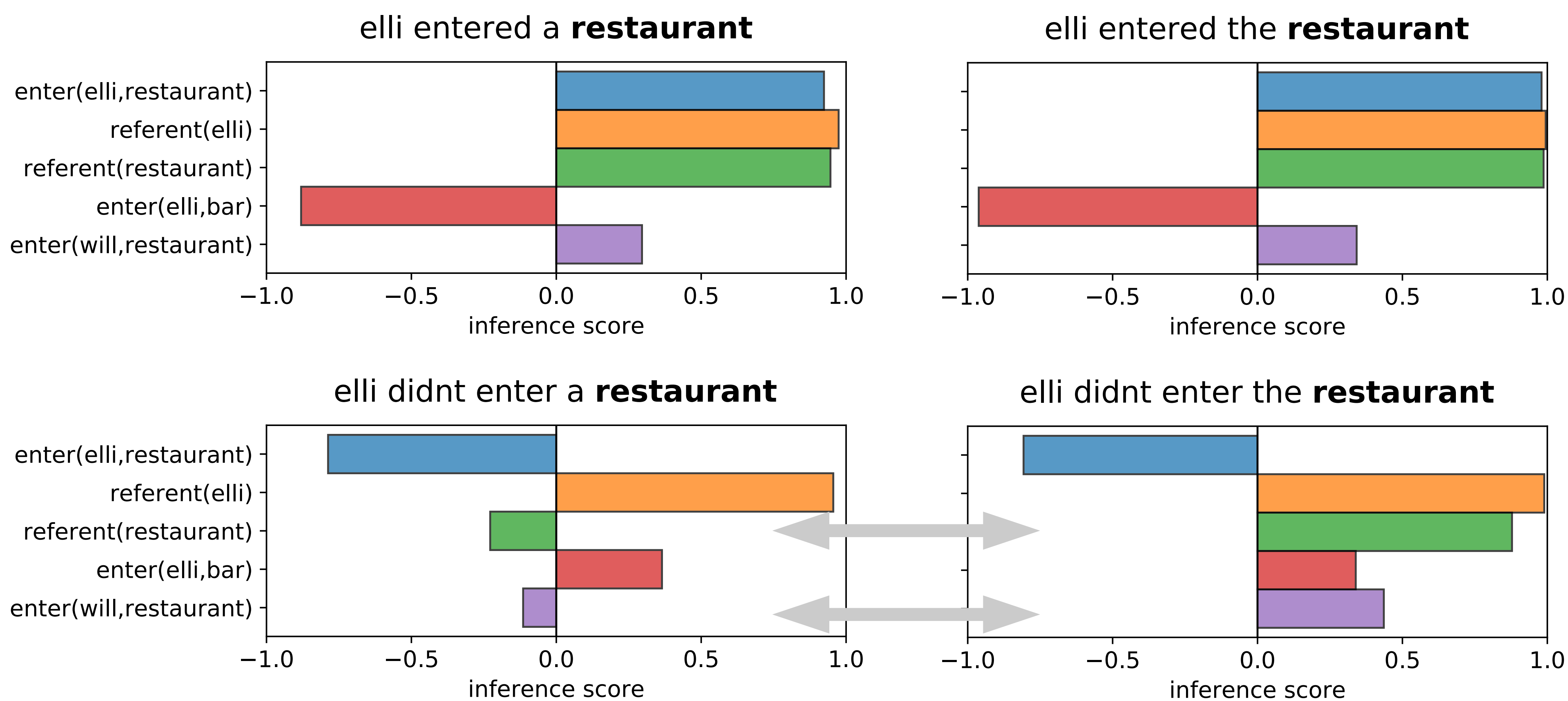
- **Surprisal** quantifies the expectancy of words in context
- Higher Surprisal: increased processing cost (Hale, 2001; Levy, 2008)
- In DFS, Surprisal quantifies expectancy of **transitions** in meaning space (from a to b), triggered by message m_{ab}:

$$S(m_{ab}) = -\log P(b|a)$$

➡ Word-by-word information effects of semantic construction

EXAMPLE: NEGATION AND PRESUPPOSITION

- Negation affects entailments and probabilistic inferences



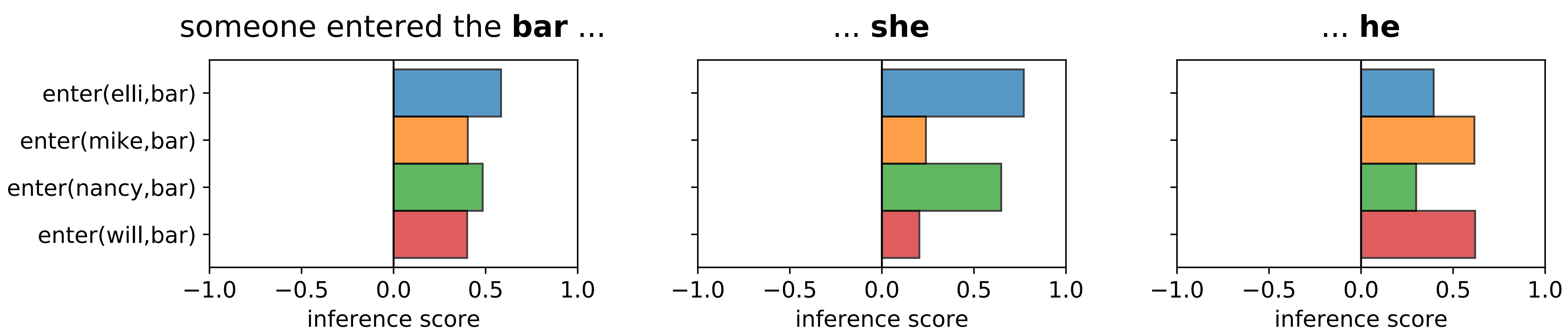
- Presupposition has an effect beyond the literal meaning

Venhuizen et al. (2022, Inf. Comp.)

EXAMPLE: QUANTIFICATION AND REFERENCE

Quantified expressions induce inferential uncertainty

- ➤ Selective expressions (e.g. pronouns) reduce this uncertainty



$S(\text{“she”}) = 0.92$ $S(\text{“he”}) = 1.12$

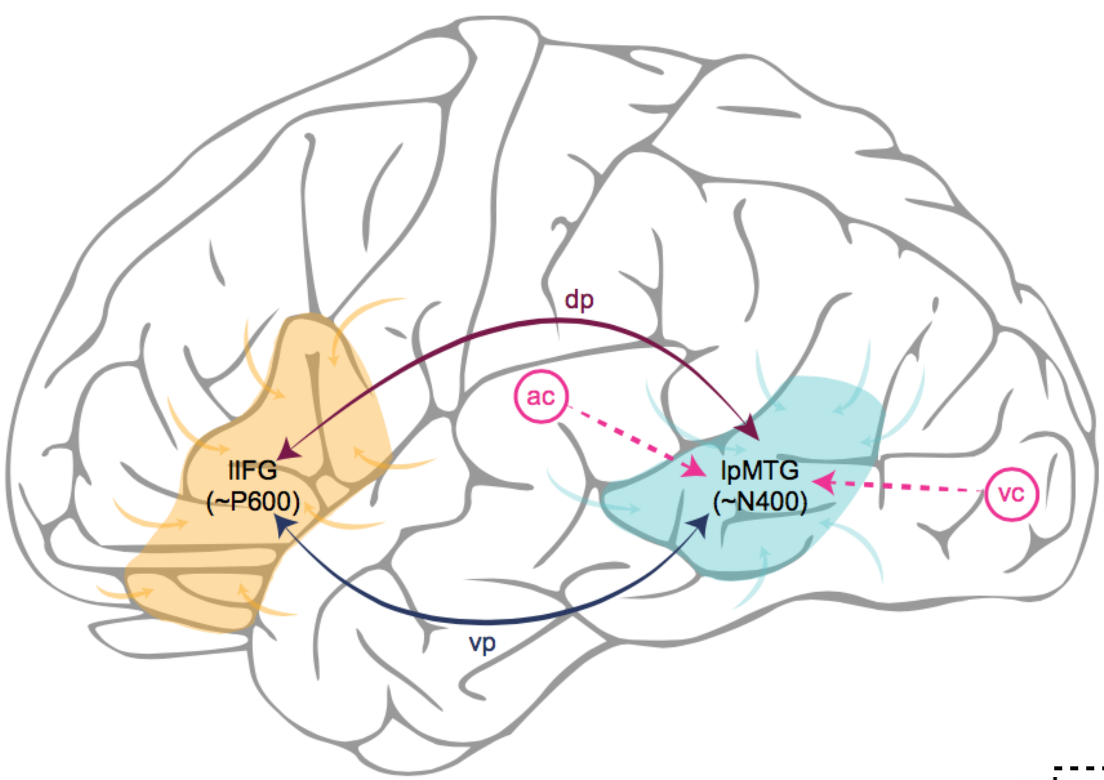
- ➤ **Surprisal** reflects expectedness of the continuation based on world knowledge and linguistic experience

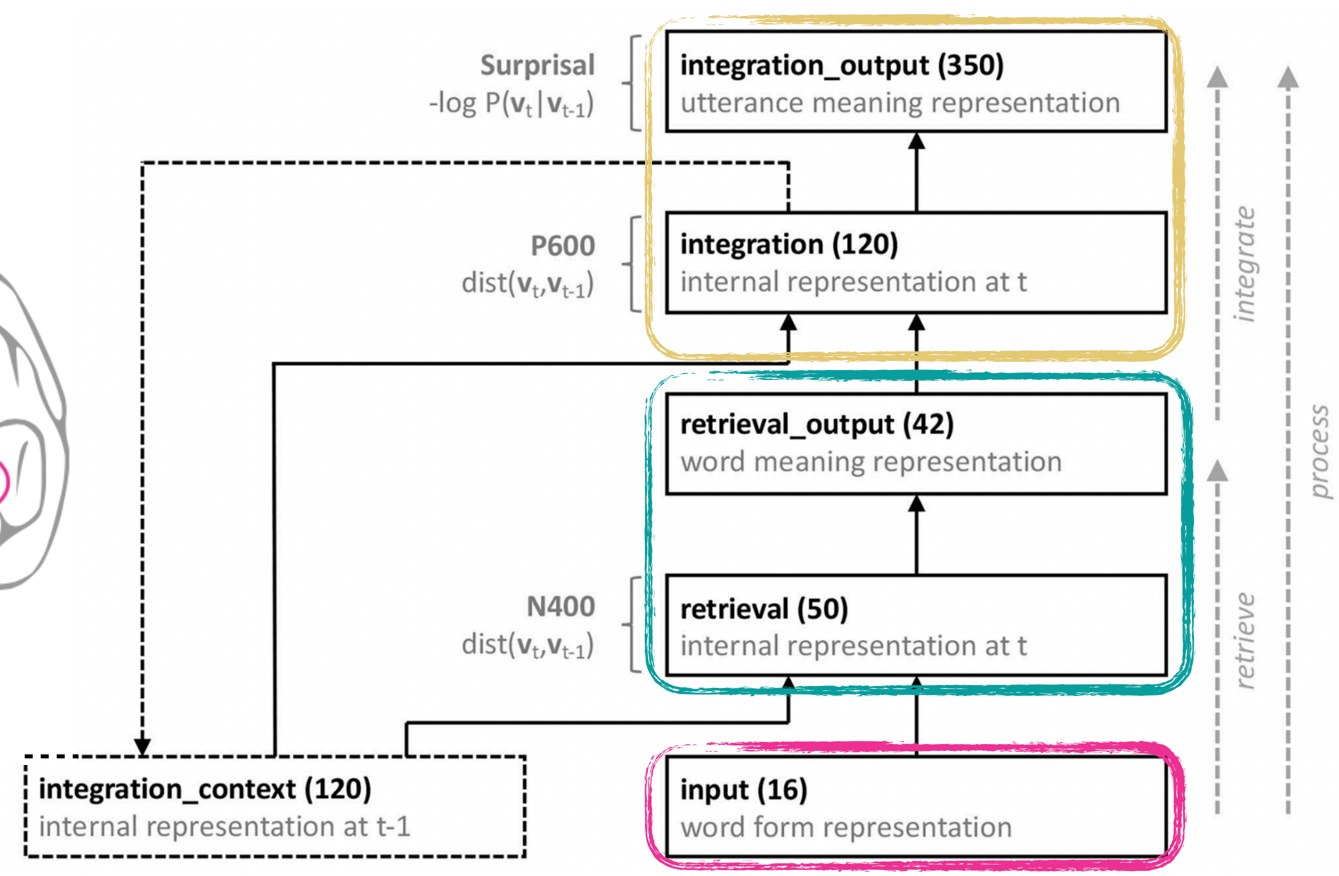
Venhuizen et al. (2022, *Inf. Comp.*)

DFS IN THE NEUROCOMPUTATION OF LANGUAGE

Word-by-word comprehension proceeds in Retrieval-Integration cycles:

- ➤ **Retrieval** of word meaning from semantic memory ~ **N400**
- ➤ **Integration** into unfolding utterance meaning ~ **P600**
- ➤ **Expectedness** of change in utterance meaning ~ **Surprisal (RT)**



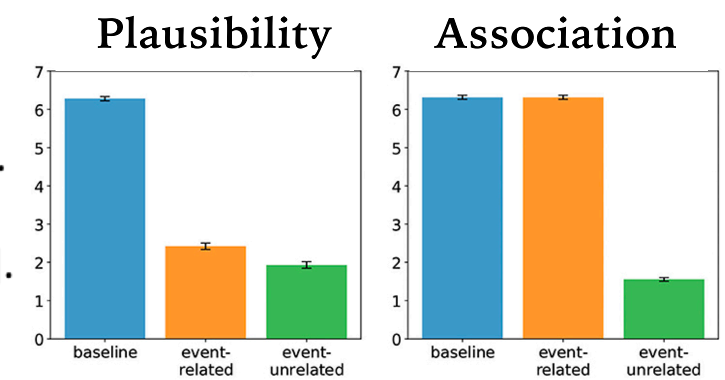


Brouwer et al. (2012, Brain Res.; 2017, Cogn. Sci.)

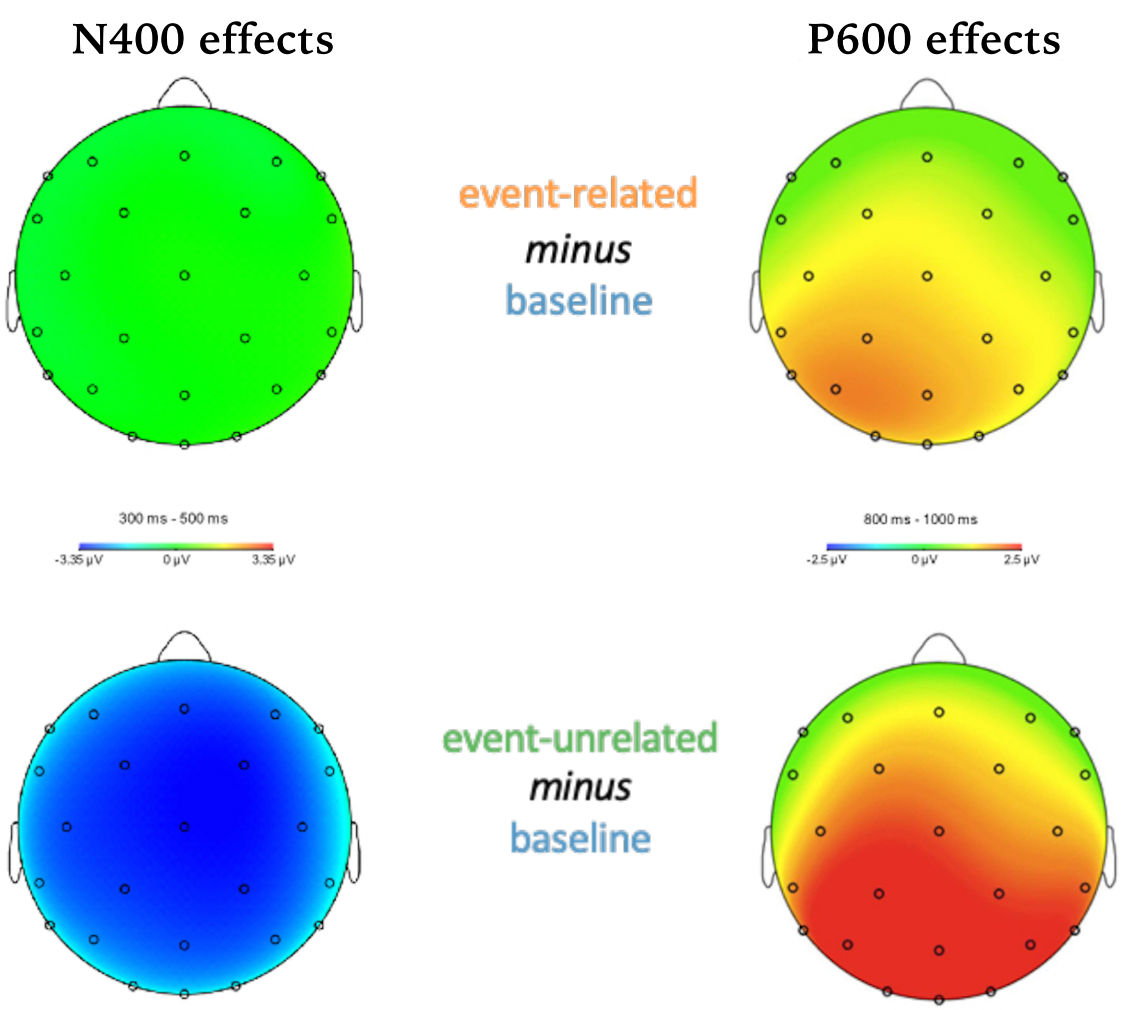
Brouwer, Delogu, Venhuizen, Crocker (2021, Frontiers Psych.)

NEURO-BEHAVIOURAL DATA: EXPERIMENTAL FINDINGS

baseline: John entered the restaurant. Before long he opened the menu [...].
event-related implausible: John left the restaurant. Before long he opened the menu [...].
event-unrelated implausible: John entered the apartment. Before long he opened the menu [...].



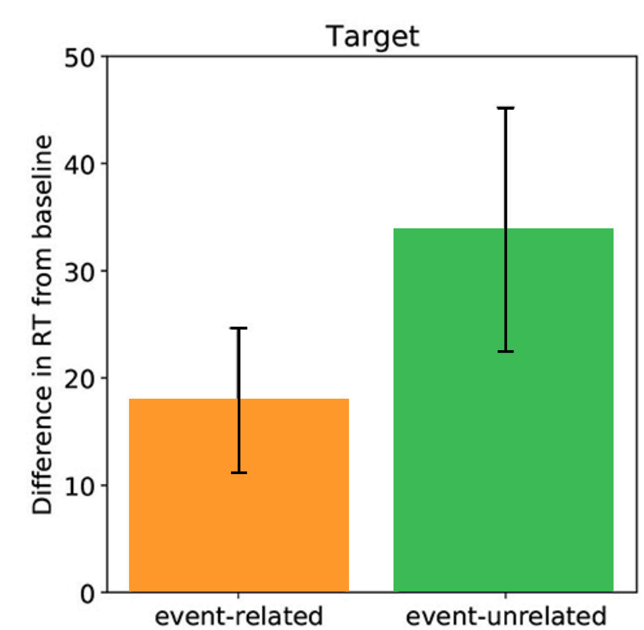
Event-related Potentials (ERPs)



plausibility only:
↑P600

plausibility & association:
↑N400
↑P600

Reading Times

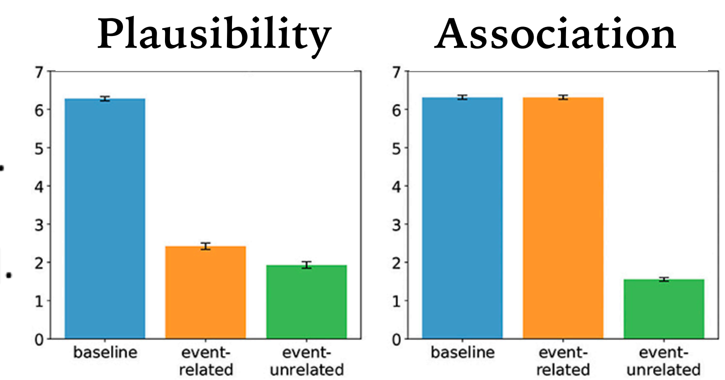


Delogu, Brouwer, Crocker (2019, Brain and Cogn.; 2021, Brain Res.)

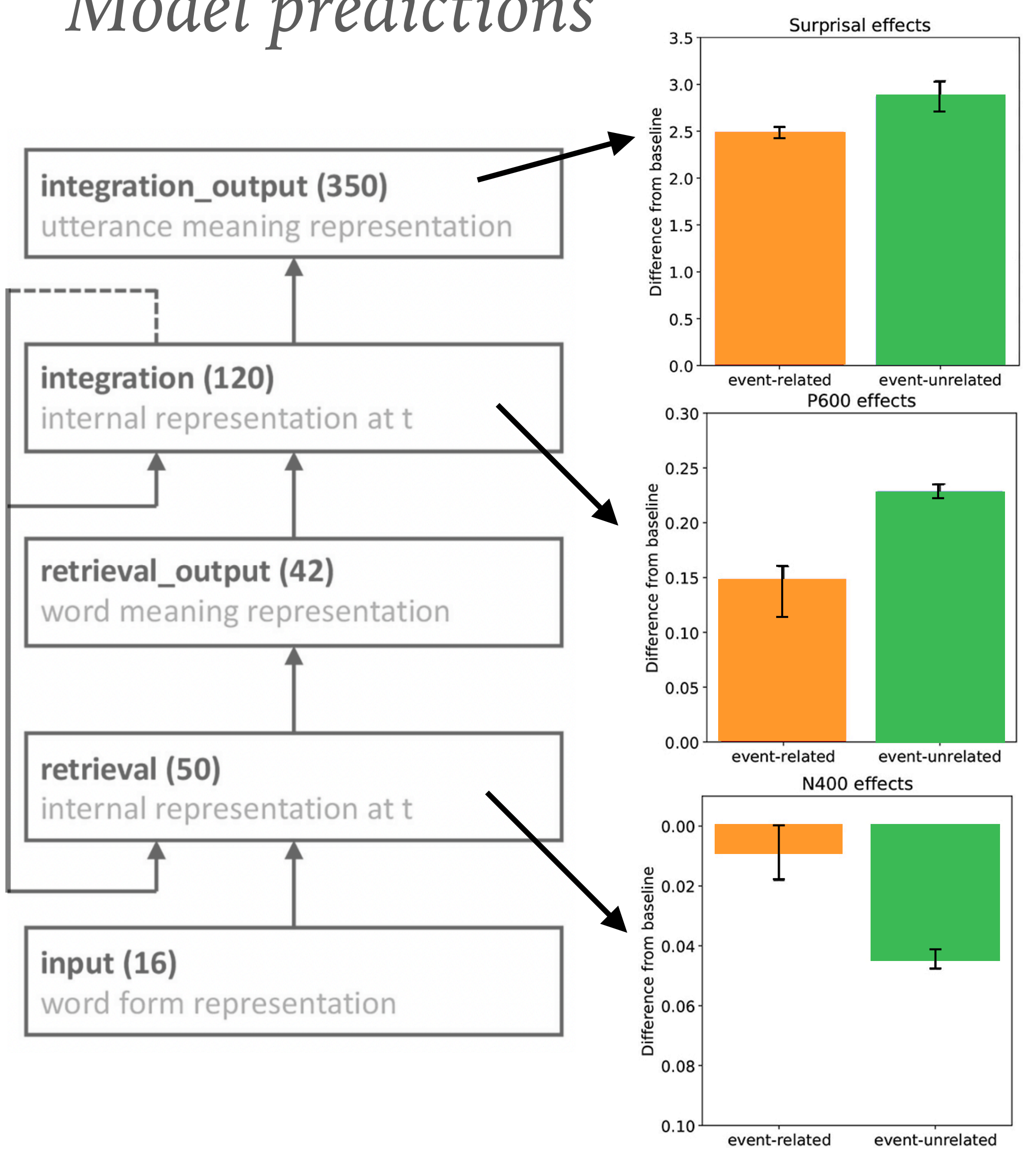
Brouwer, Delogu, Venhuizen, Crocker (2021, Frontiers Psych.)

NEURO-BEHAVIOURAL DATA: MODEL PREDICTIONS

baseline: John entered the restaurant. Before long he opened the menu [...].
event-related implausible: John left the restaurant. Before long he opened the menu [...].
event-unrelated implausible: John entered the apartment. Before long he opened the menu [...].



Model predictions



↑Surprisal: expectancy

↑P600: integration

↑N400: retrieval

Surprisal ~ RT		
Condition	**Data**	**Model**
Baseline	—	—
Event-rel.	Yes	✓
Event-unrel.	Yes	✓

P600		
Condition	**Data**	**Model**
Baseline	—	—
Event-rel.	Yes	✓
Event-unrel.	Yes	✓

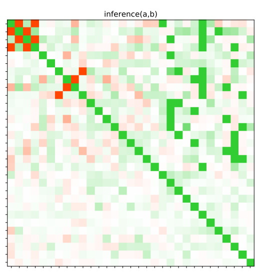
N400		
Condition	**Data**	**Model**
Baseline	—	—
Event-rel.	No	✓
Event-unrel.	Yes	✓

Brouwer, Delogu, Venhuizen, Crocker (2021, Frontiers Psych.)

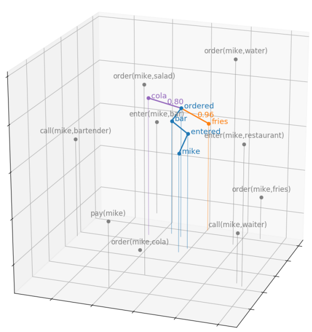
EXPECTATION-BASED SEMANTICS IN LANGUAGE COMPREHENSION

	p_1	p_2	p_3	…
M_1	1	1	0	…
M_2	1	0	0	…
M_3	0	1	0	…
M_4	1	1	1	…
…	…	…	…	…

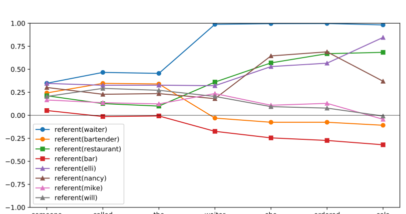
Meaning space from proposition-level co-occurrences



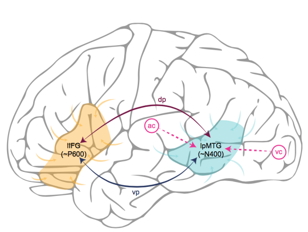
Compositional semantics, entailment and probabilistic inferences using vector operations



Dynamic meaning construction as word-by-word navigation through the meaning space



Expectations from proposition-level world knowledge and linguistic experience

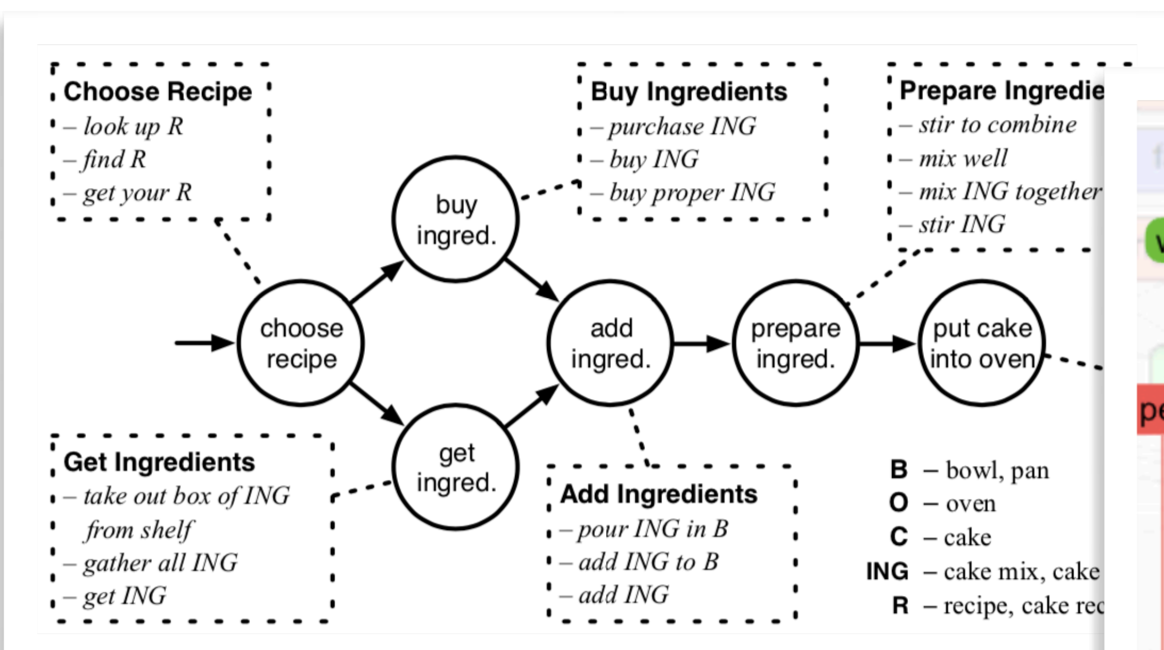


Division of labour between lexical- and utterance-level semantics in modeling language comprehension

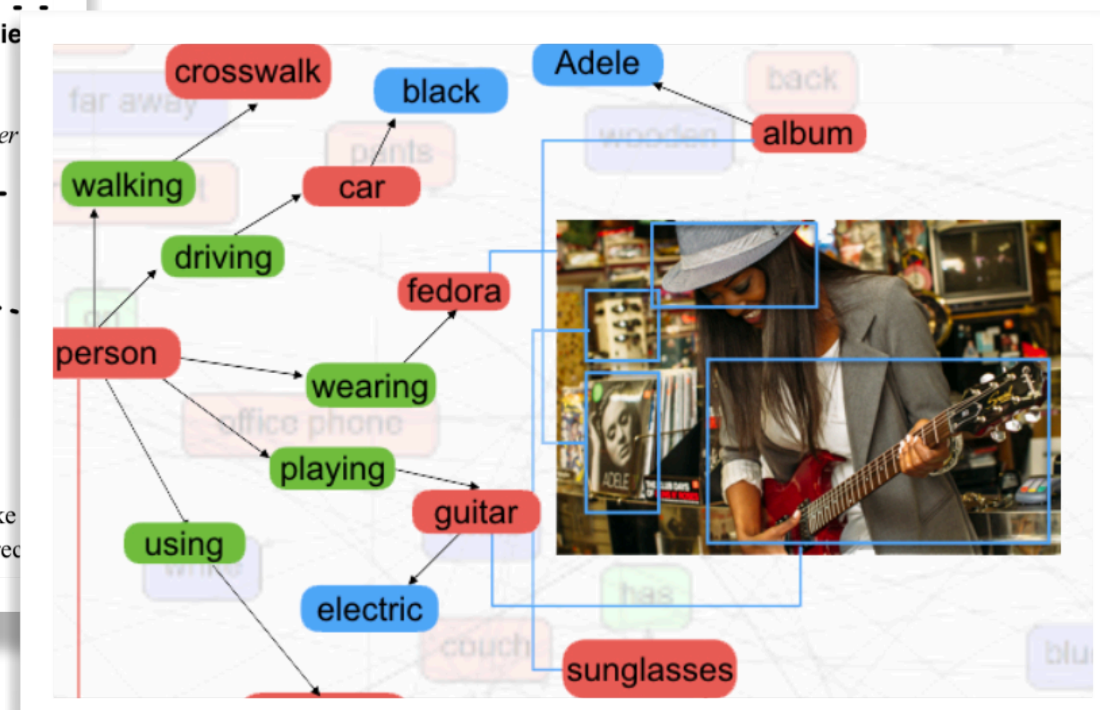
OUTLOOK: DATA-DRIVEN DFS

Goal: Employ DFS to model deep understanding using real-world propositional co-occurrences

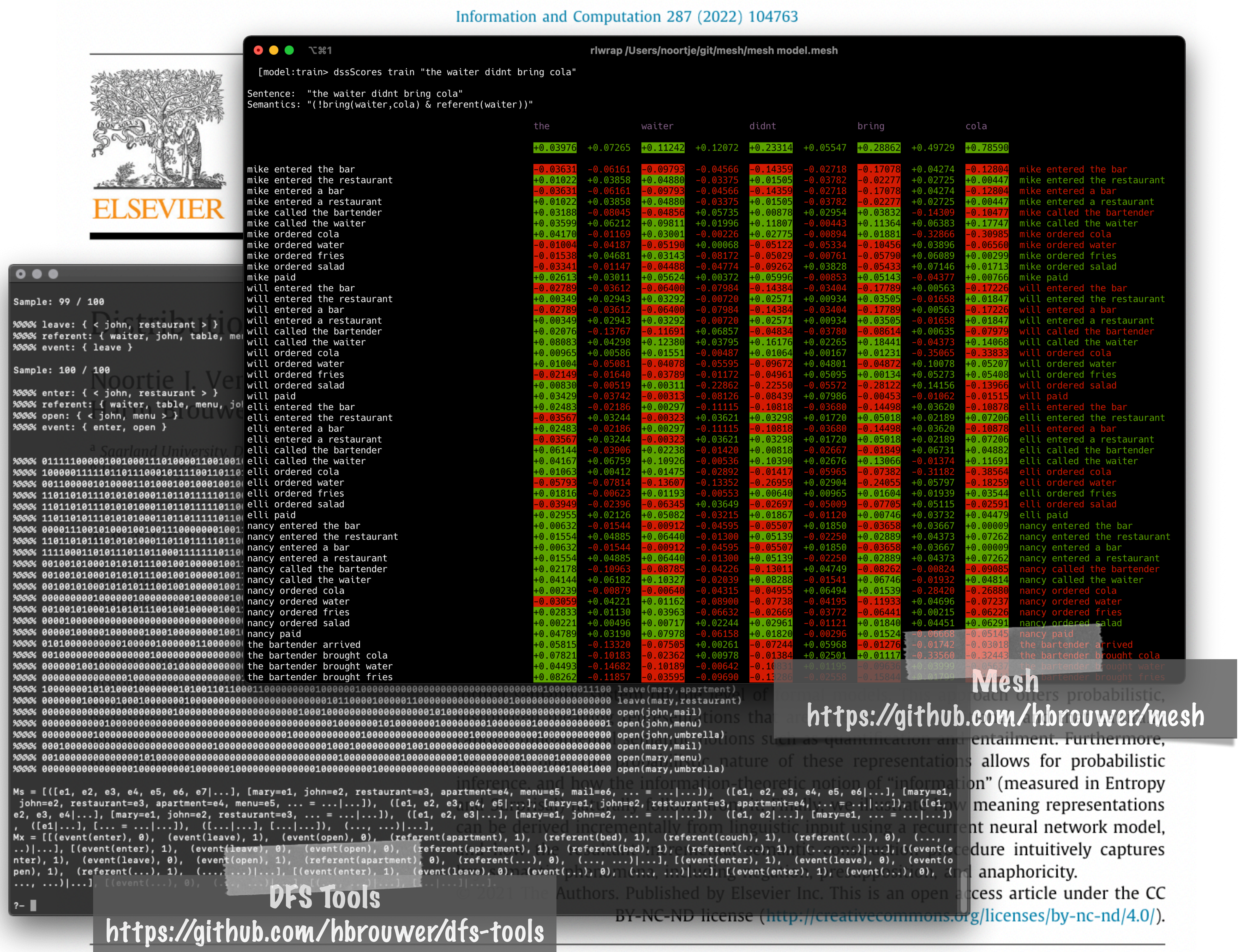
- **Step 1**: Derive meaning space from resource with world knowledge about propositional (not: lexical) co-occurrence
- **Step 2**: Mapping natural language to vectors in the meaning space



DeScript corpus (Wanzare et al., 2016)



Visual Genome (Krishna et al., 2016)



Information and Computation 287 (2022) 104763
ELSEVIER
rlwrap /Users/noortje/git/mesh/mesh model.mesh
[model:train> dssScores train "the waiter didnt bring cola"
Sentence: "the waiter didnt bring cola"
Semantics: "(!bring(waiter,cola) & referent(waiter))"
the
waiter
didnt
bring
cola
Sample: 99 / 100
%%%% leave: { < john, restaurant > }
%%%% event: { leave }
Sample: 100 / 100
%%%% enter: { < john, restaurant > }
%%%% open: { < john, menu > }
%%%% event: { enter, open }
Mesh
https://github.com/hbrouwer/mesh
DFS Tools
https://github.com/hbrouwer/dfs-tools