Distributional Formal Semantics

Noortje Venhuizen

Petra Hendriks
Matthew Crocker
Harm Brouwer

NATURAL LANGUAGE SEMANTICS

Model-theoretic Semantics

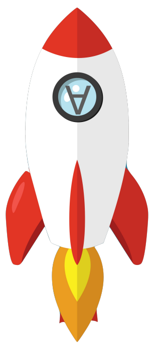
- Truth-conditional meaning
- Logical entailment
- Compositionality

Distributional Semantics

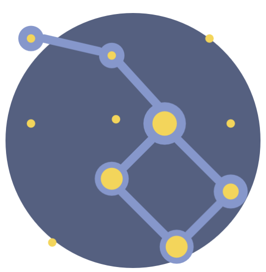
- Semantic similarity
- Empirically driven
- Cognitively inspired

E.g., Baroni et al. (2010,2014); Boleda & Herbelot (2016); Coecke et al. (2010); Grefenstette & Sadrzadeh (2011); Socher et al. (2012)

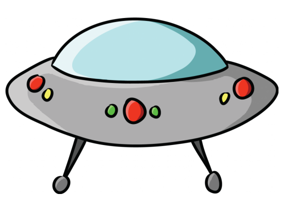
A FRAMEWORK FOR DISTRIBUTIONAL FORMAL SEMANTICS



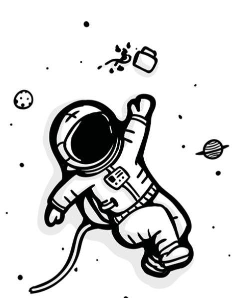
A meaning space for Distributional Formal Semantics



Formal properties of the meaning space

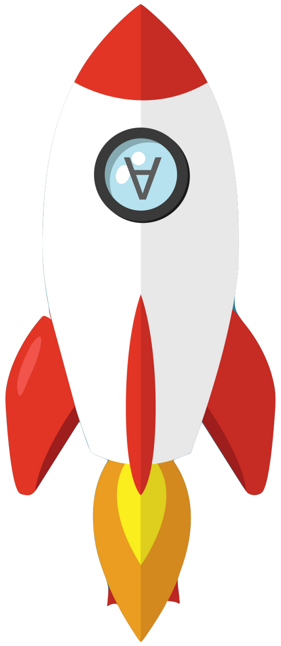


Incremental meaning construction



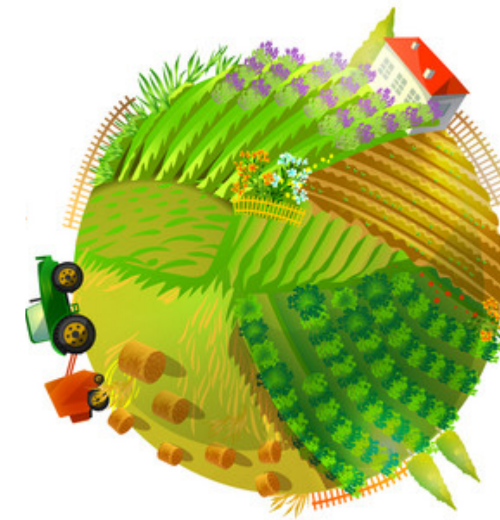
Semantic processing in the meaning space

A MEANING SPACE FOR DISTRIBUTIONAL FORMAL SEMANTICS



FROM MODELS TO MEANING SPACE

…



$M_1 = \langle U_1, V_1 \rangle$ $M_2 = \langle U_2, V_2 \rangle$ $M_3 = \langle U_3, V_3 \rangle$ $M_n = \langle U_n, V_n \rangle$

$p_1 \wedge \neg p_2 \wedge p_3 \wedge \ldots$ $\neg p_1 \wedge p_2 \wedge p_3 \wedge \ldots$ $\neg p_1 \wedge p_2 \wedge \neg p_3 \wedge \ldots$ $\neg p_1 \wedge \neg p_2 \wedge \neg p_3 \wedge \ldots$

- The set of models $\mathcal{M}_{\mathcal{P}}$ — describing states-of-affairs over propositions in $\mathcal{P}$ — defines a meaning space
- Propositional meaning defined by co-occurrence across models

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2021, Inf. Comp./ArXiv)

CAPTURING THE STRUCTURE OF THE WORLD

"A boy rides a bike"

Boy is (likely) outside

Boy is not asleep

If it's evening, the light is on

The bike has wheels

etc.

World knowledge restricts propositional co-occurrence in the meaning space derived from the set of models $\mathcal{M}_\mathcal{P}$

- ➤ **Hard** world knowledge constraints restrict individual models
- ➤ **Probabilistic** constraints define probabilistic co-occurrences across the set of models $\mathcal{M}_\mathcal{P}$

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2021, Inf. Comp./ArXiv)

DFS MEANING SPACE $S_{\mathcal{M}\times\mathcal{P}}$

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	…
…	…	…	…	…	…

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

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

- **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

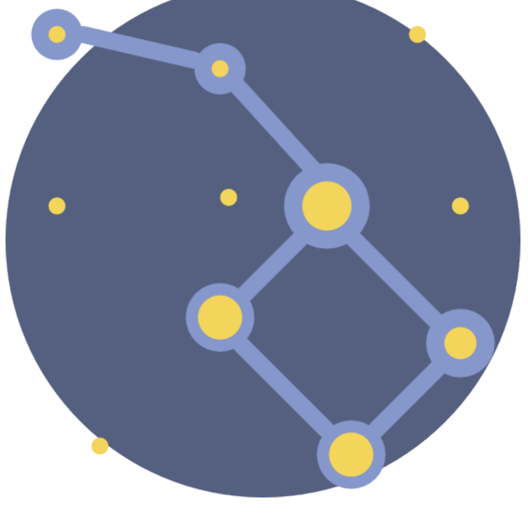
THE DISTRIBUTIONAL HYPOTHESIS REVISITED

“

You shall know a ~~word~~ *proposition* by the company it keeps

- J. R. Firth (1957)

FORMAL PROPERTIES OF THE MEANING SPACE



MEANING VECTOR COMPOSITION

Meaning vectors can be combined to define compositional meanings

- Standard logical operators interpreted as in model-theory

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

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

… 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 $\mathcal{M}_\mathcal{P}$)

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

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

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2021, Inf. Comp./ArXiv)

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	…
	…	…	…	…	…

Venhuizen, Hendriks, Crocker, Brouwer (2019, WoLLIC; 2021, Inf. Comp./ArXiv)

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)

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

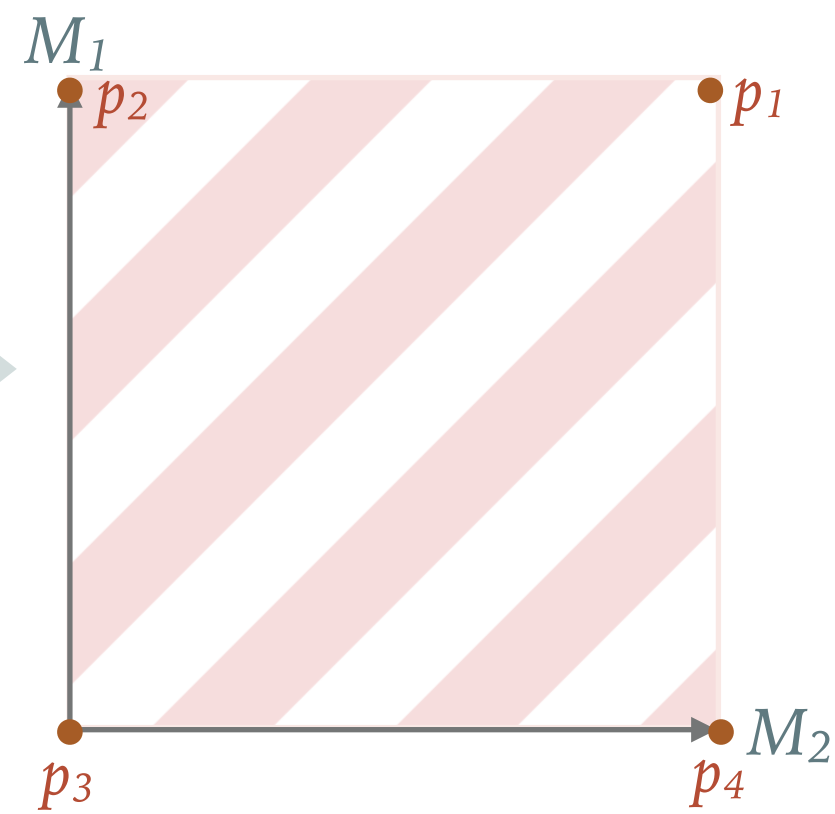
- $P(a|b) \leq P(a)$: Negative inference (b decreases probability of a)

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

Frank et al. (2009,Cognition)

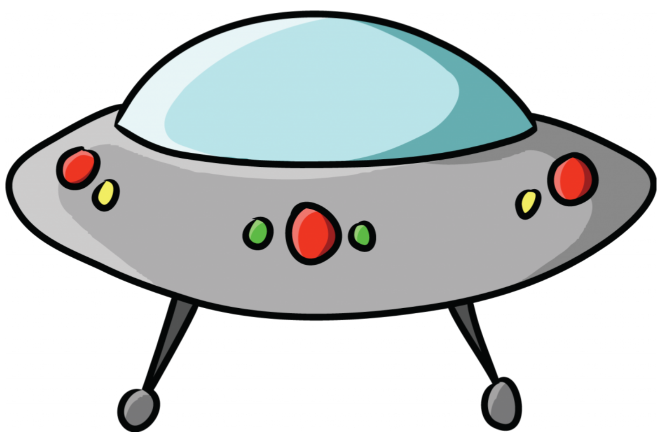
CONTINUOUS NATURE OF THE MEANING SPACE

	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

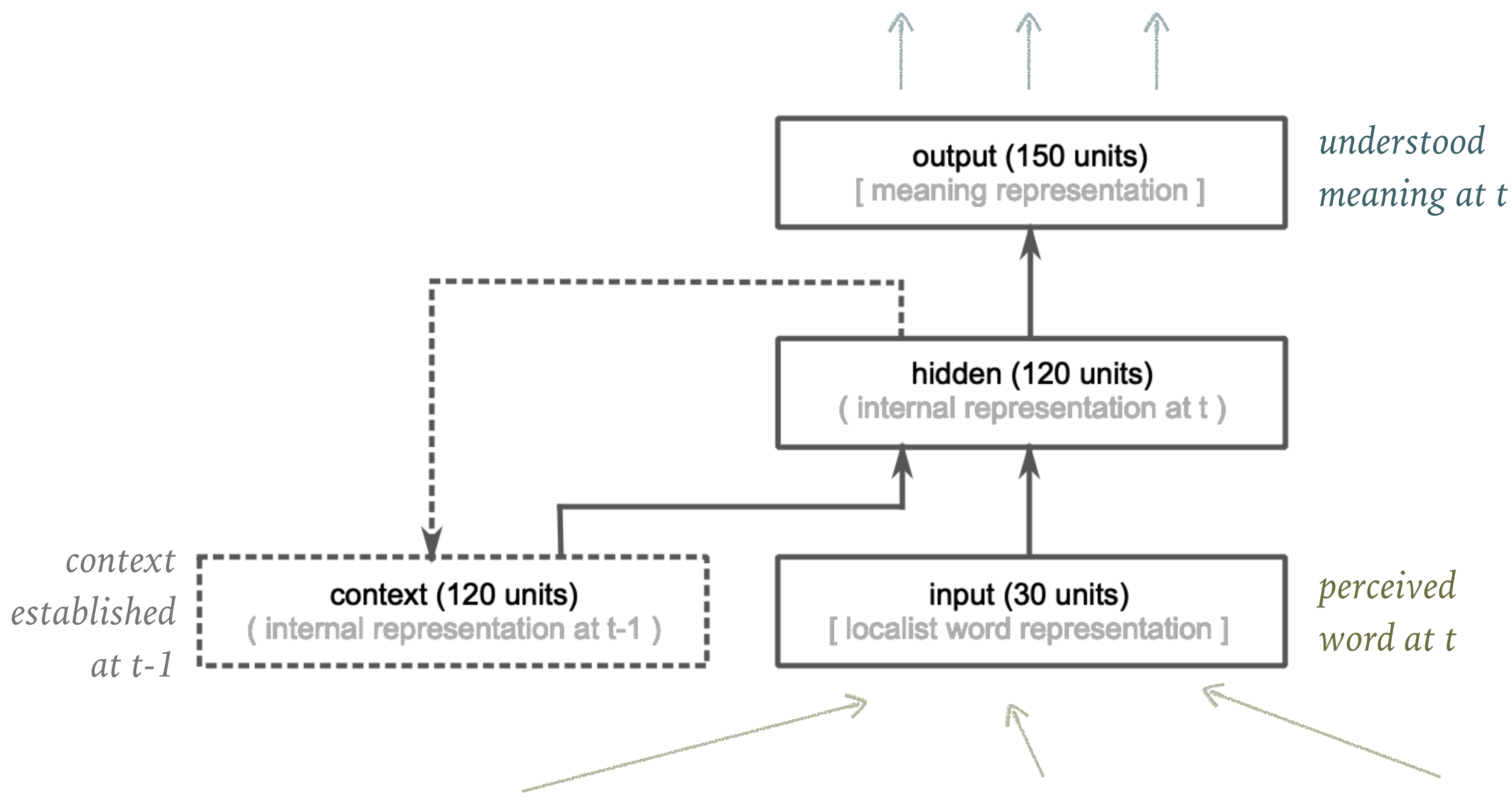
INCREMENTAL MEANING CONSTRUCTION



A MODEL OF INCREMENTAL MEANING CONSTRUCTION

enter(mike,restaurant) ∧ order(mike,salad)

[1,0,1,0,0,1,1,0,1,0,0,1,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,…]



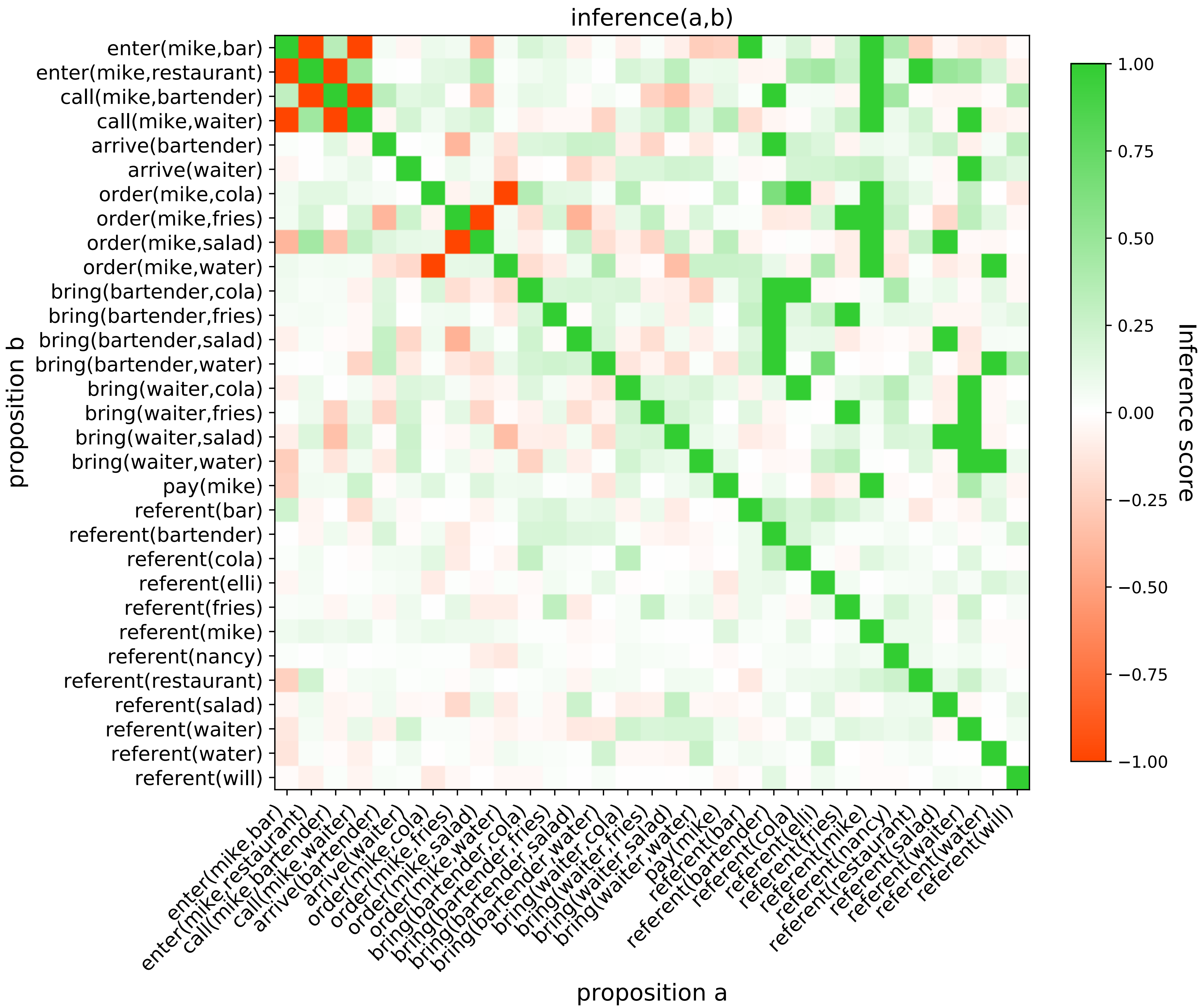
< [0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0], [0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0], […] >

<“mike”, “entered”, “the”, “restaurant”, “and” “ordered”, “salad”>

“Mike entered the restaurant and ordered salad”

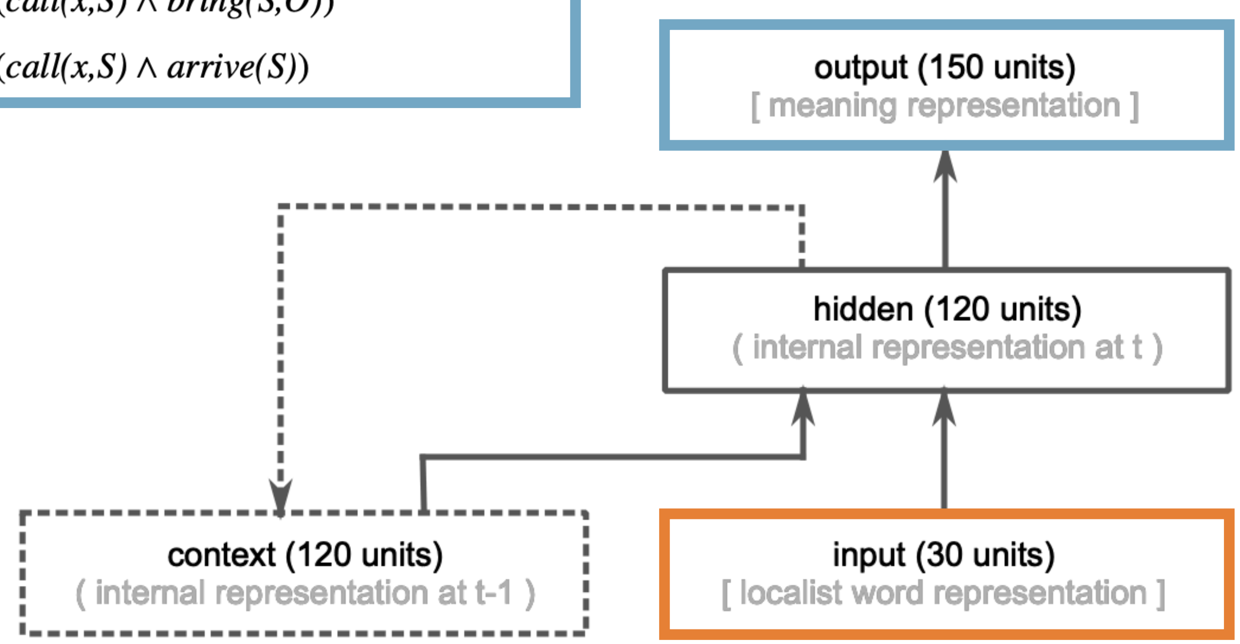
CONSTRUCTING THE MODEL: MEANING SPACE

We sampled a meaning space of 150 models describing 51 propositions

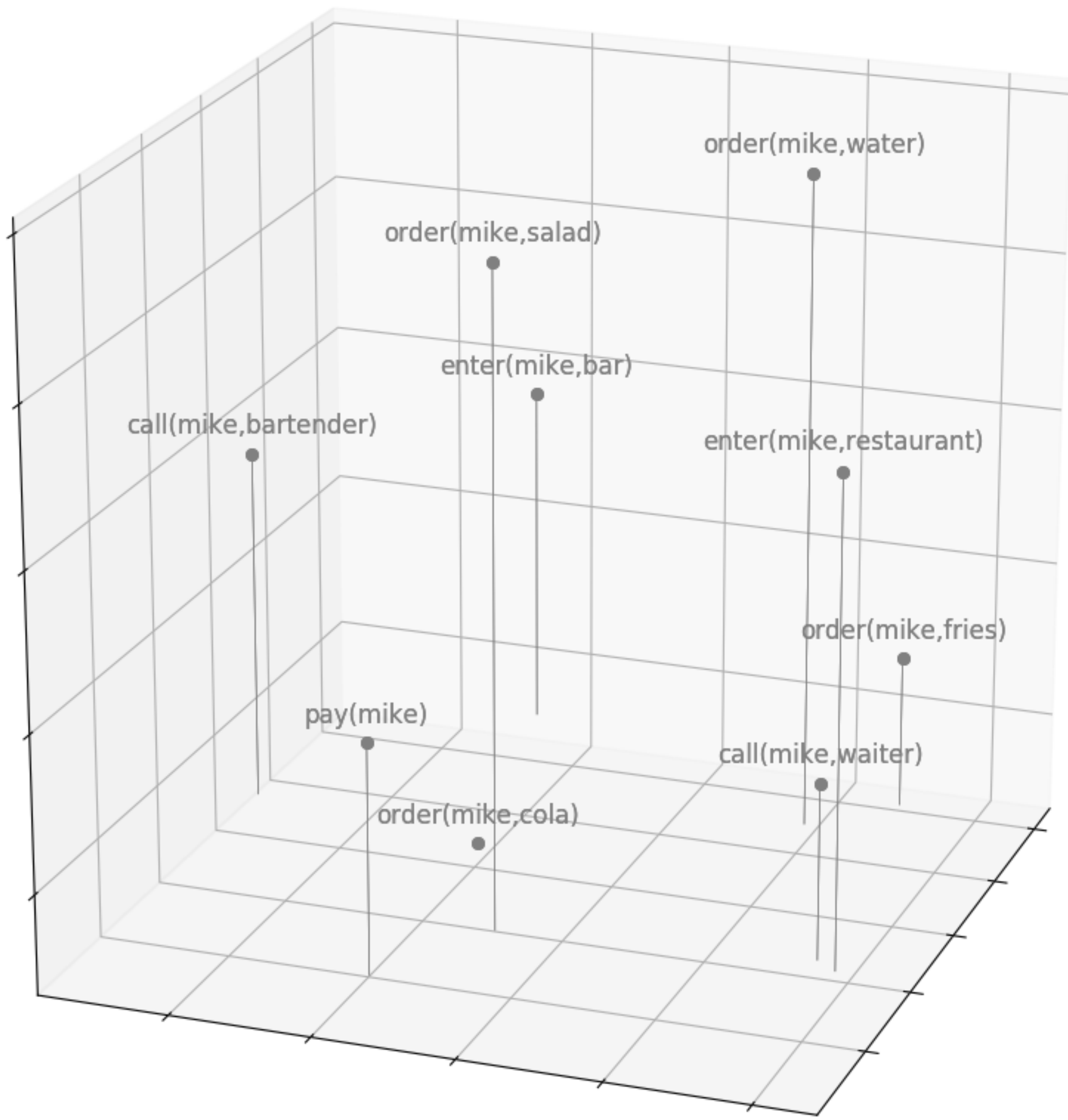


CONSTRUCTING THE MODEL: LANGUAGE

Utterance$_{[a/b]}$	**Semantics**$_a$	**Semantics**$_b$
P [entered/didn't enter] a L	$enter(P,L)$	$\neg(enter(P,L)) \wedge referent(P)$
P [entered/didn't enter] the L	$enter(P,L)$	$\neg(enter(P,L)) \wedge referent(P) \wedge referent(L)$
P [called/didn't call] the S	$call(P,S)$	$\neg(call(P,S)) \wedge referent(P) \wedge referent(S)$
P [ordered/didn't order] O	$order(P,O)$	$\neg(order(P,O)) \wedge referent(P)$
P [paid/didn't pay]	$pay(P)$	$\neg(pay(P)) \wedge referent(P)$
the S [arrived/didn't arrive]	$arrive(S)$	$\neg(arrive(S)) \wedge referent(S)$
the S [brought/didn't bring] O	$bring(S,O)$	$\neg(bring(S,O)) \wedge referent(S)$
[P/someone] entered the L he/she ordered O	$enter(P,L) \wedge order(P,O)$	$\exists x_{m/f}(enter(x,L) \wedge order(x,O))$
[P/someone] entered the L he/she called the S	$enter(P,L) \wedge call(P,S)$	$\exists x_{m/f}(enter(x,L) \wedge call(x,S))$
[P/someone] called the S he/she ordered O	$call(P,S) \wedge order(P,O)$	$\exists x_{m/f}(call(x,S) \wedge order(x,O))$
[P/someone] called the S he/she paid	$call(P,S) \wedge pay(P)$	$\exists x_{m/f}(call(x,S) \wedge pay(x))$
[P/someone] called the S he brought O	$call(P,S) \wedge bring(S,O)$	$\exists x(call(x,S) \wedge bring(S,O))$
[P/someone] called the S he arrived	$call(P,S) \wedge arrive(S)$	$\exists x(call(x,S) \wedge arrive(S))$

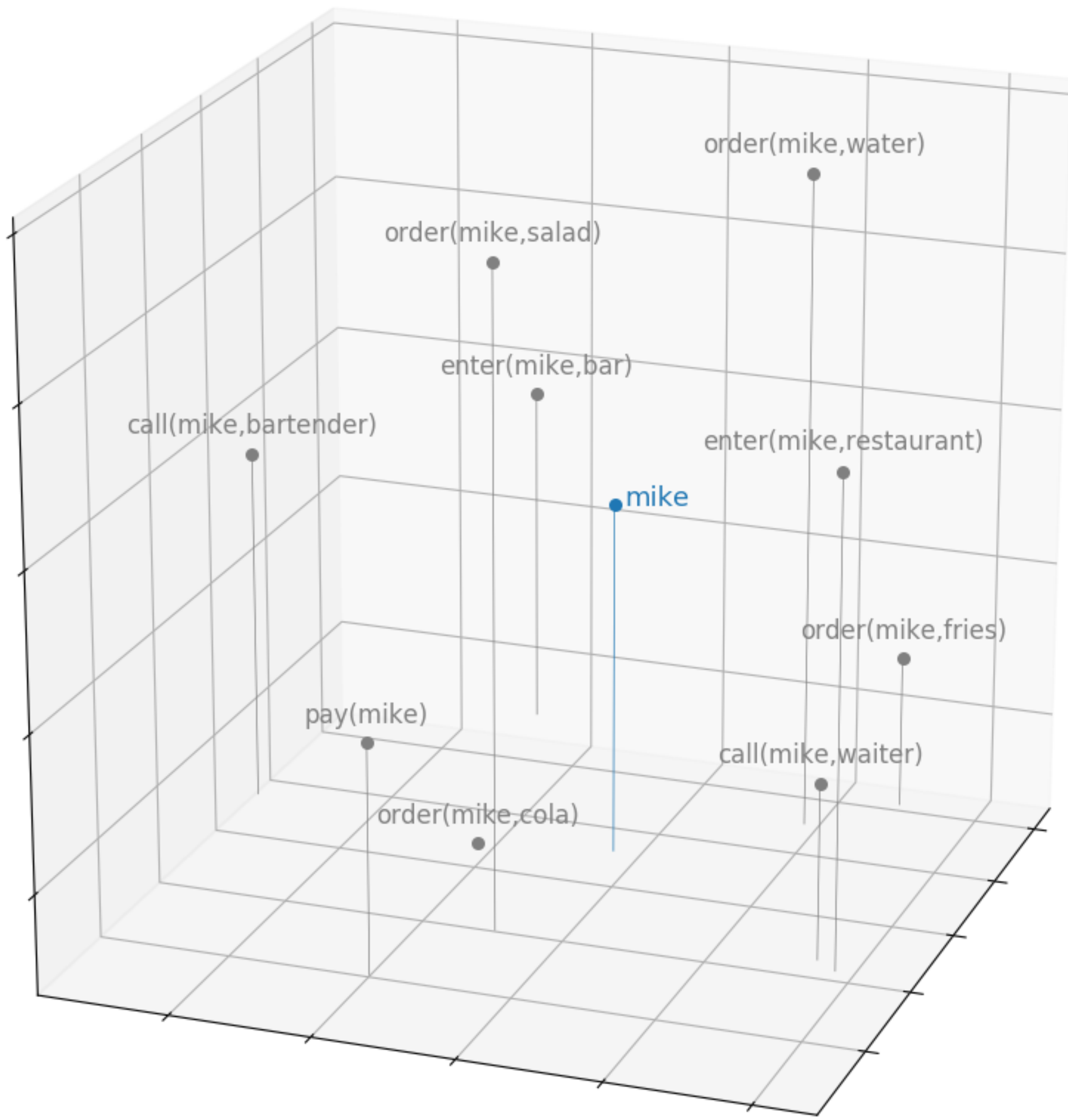


MEANING SPACE NAVIGATION



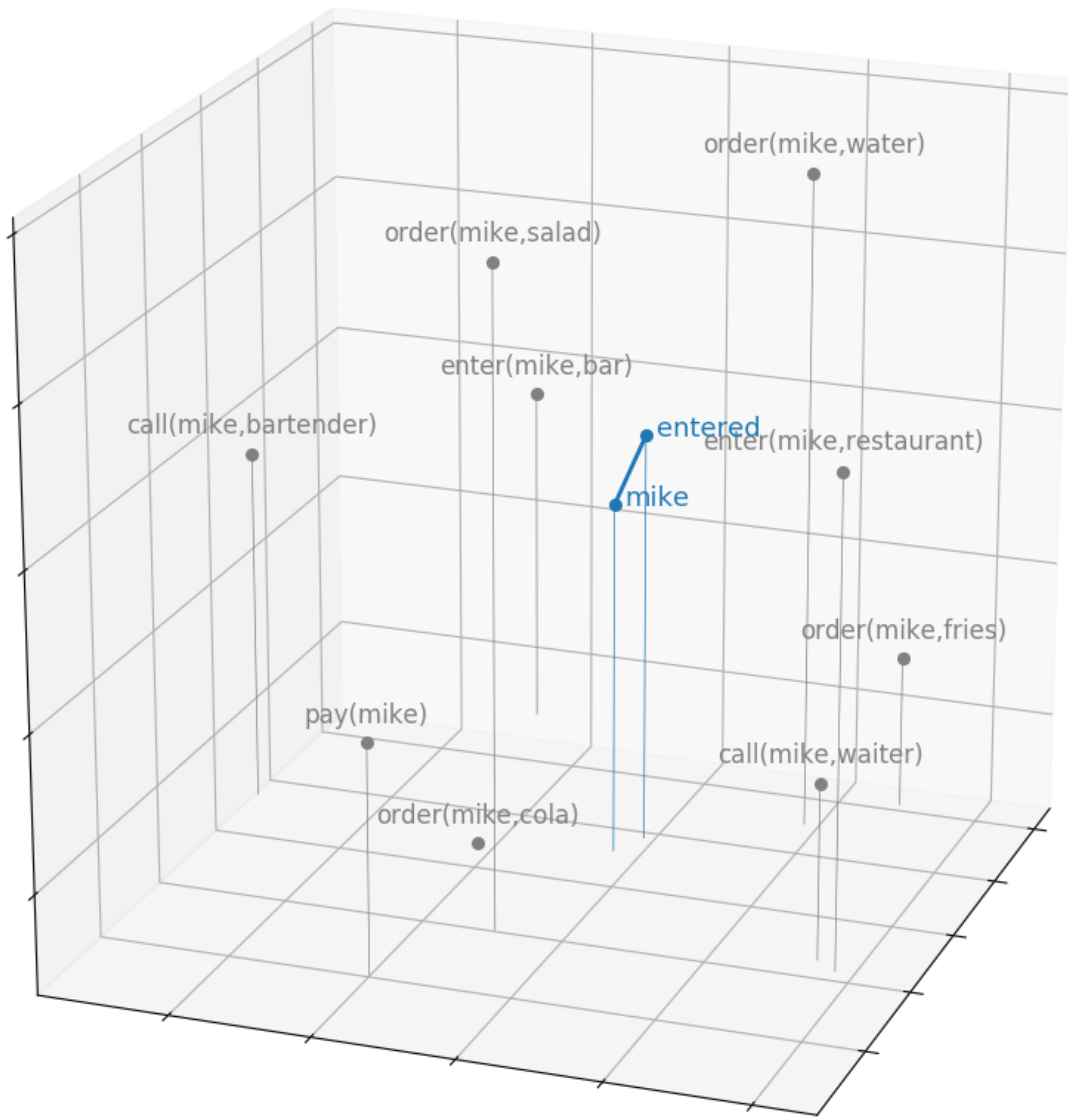
- Propositions that co-occur frequently in $\mathcal{M}$ are positioned close to each other in space

MEANING SPACE NAVIGATION



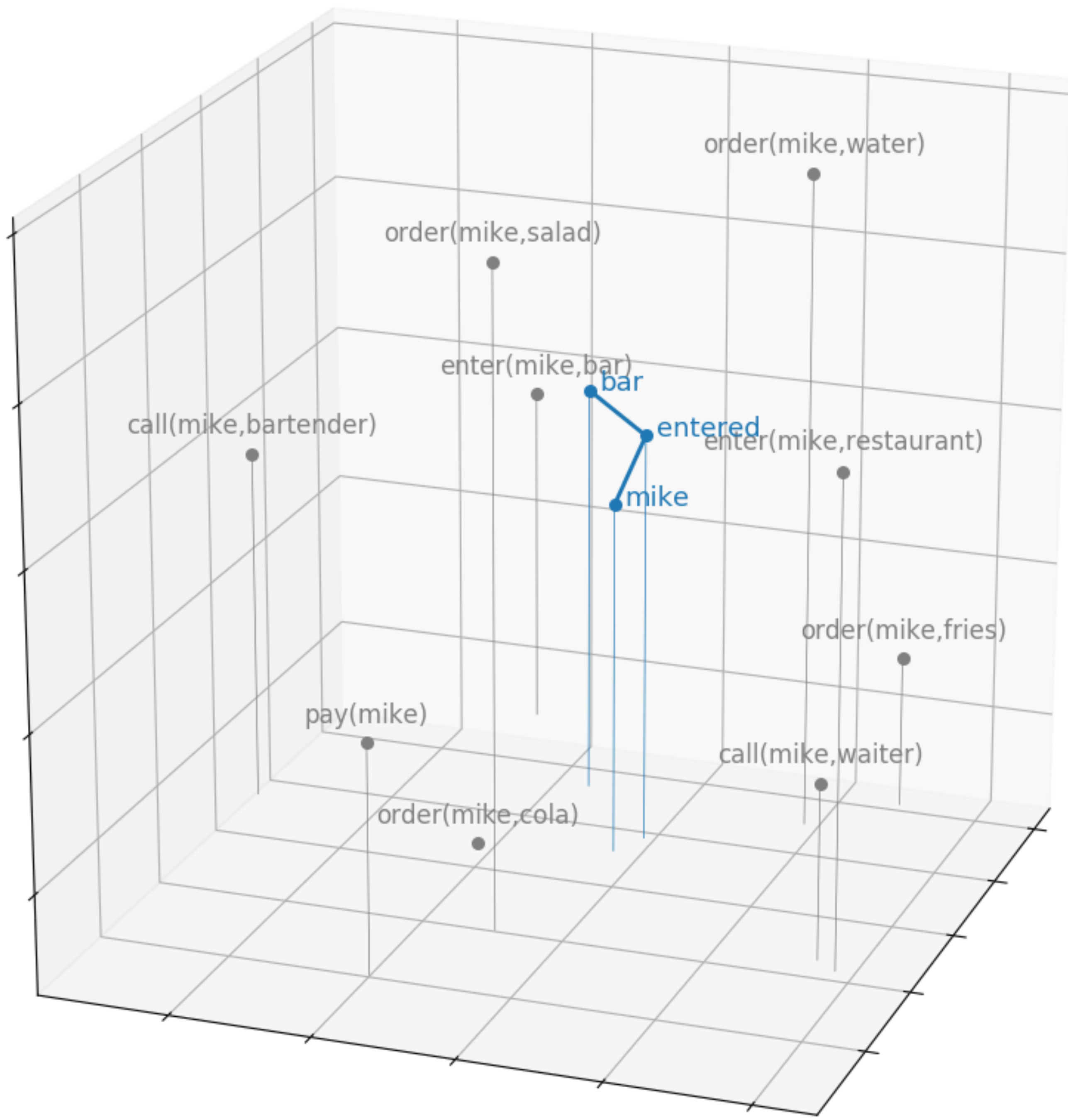
- Model-derived meaning of '*mike*' abstracts over the meanings of all propositions pertaining to *mike*

MEANING SPACE NAVIGATION



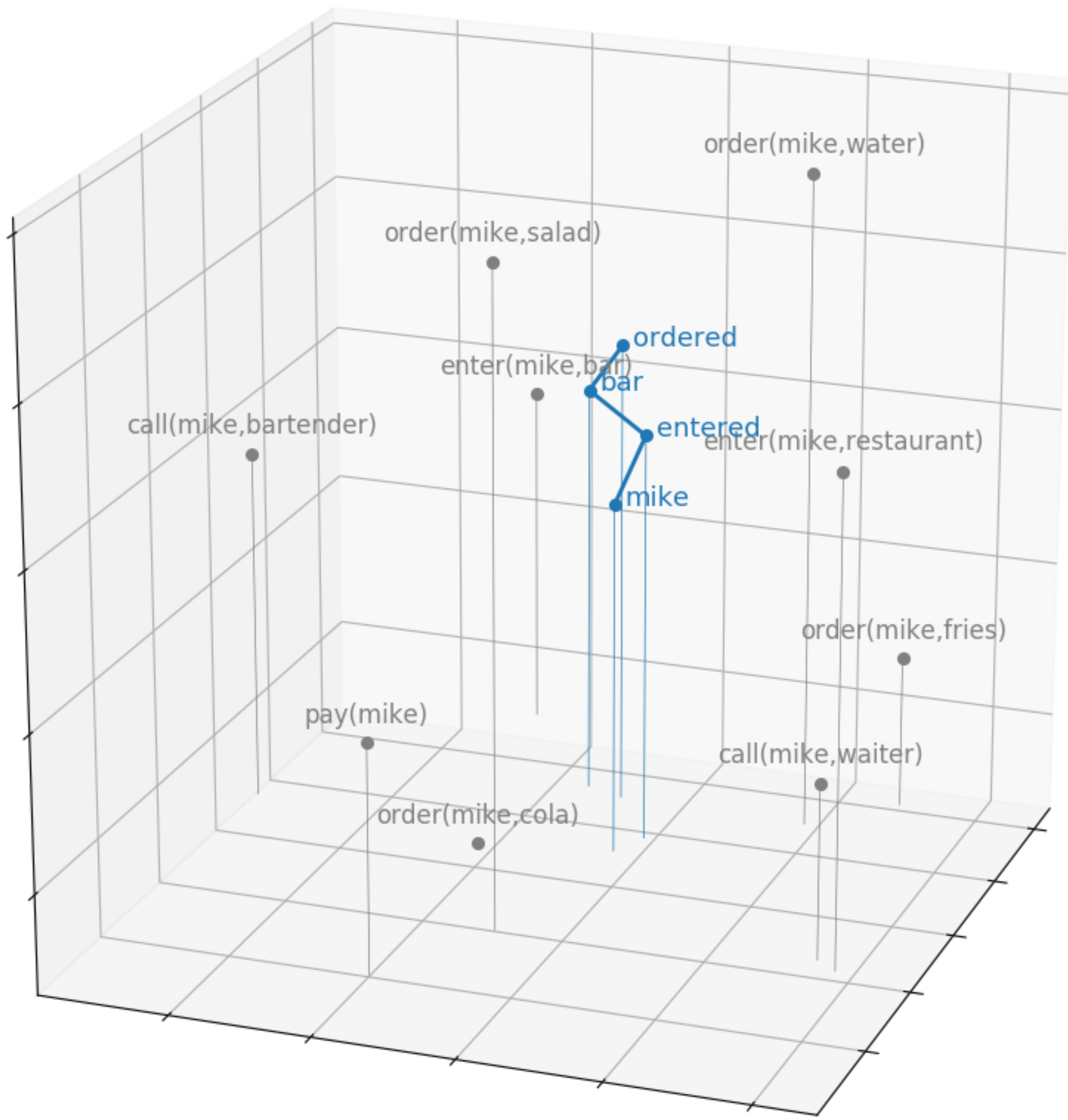
- At the word “entered”, the model navigates to a point that represents the contextualised meaning “mike entered”

MEANING SPACE NAVIGATION



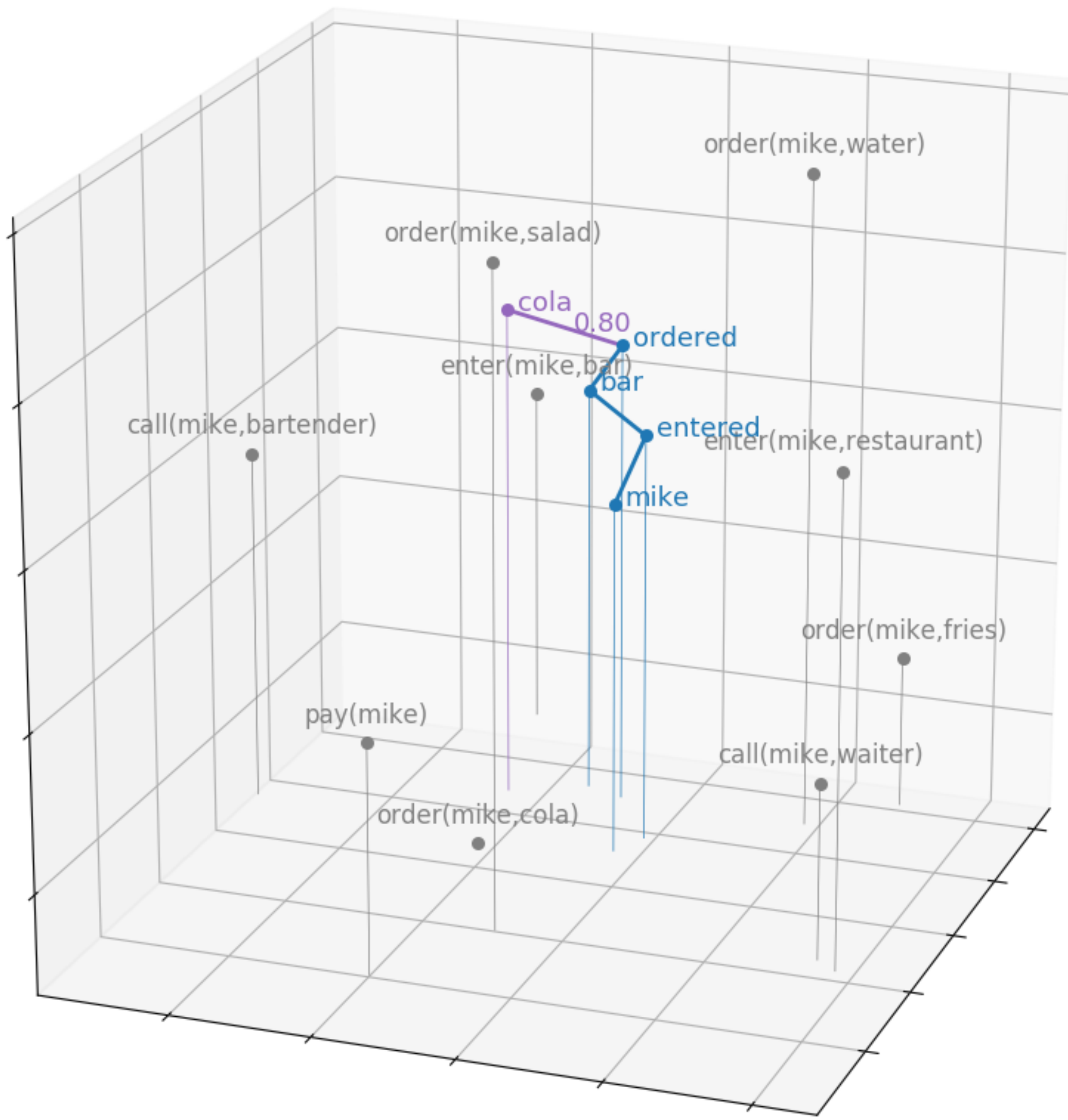
- The utterance "*mike entered the bar*" approximates the propositional meaning vector for *enter(mike,bar)*

MEANING SPACE NAVIGATION



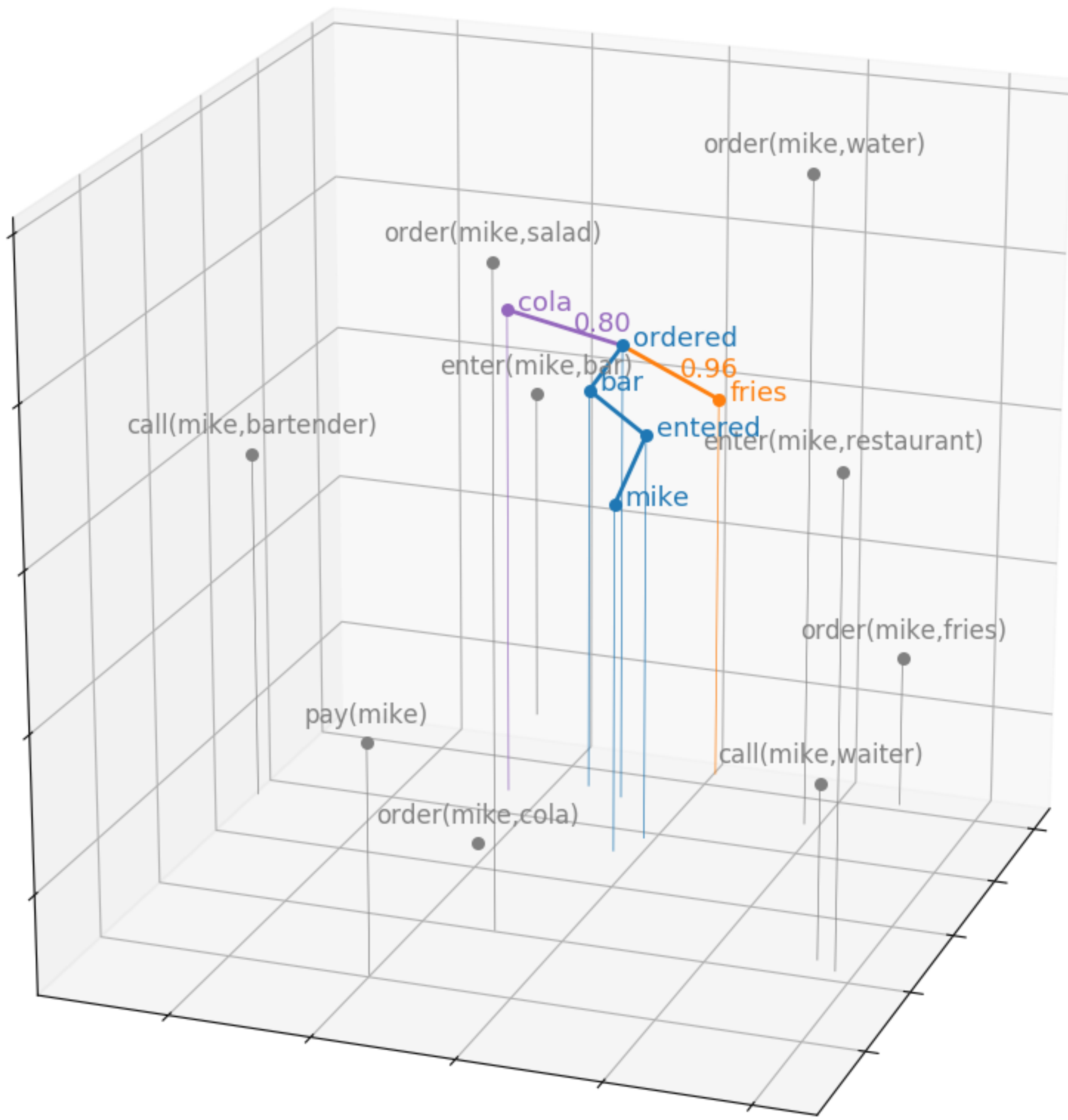
- The meaning vector after processing "*mike entered the bar [.] he ordered*" is close to *order* propositions that are typical given *enter(mike,bar)*

MEANING SPACE NAVIGATION



- When the utterance is continued with *"cola"*, the model approximates the conjunctive meaning vector *enter(mike,bar)* ∧ *order(mike,cola)*

MEANING SPACE NAVIGATION



- By contrast, *"fries"* results in different transition in meaning space—approximating the conjunctive meaning *enter(mike,bar)* ∧ *order(mike,fries)*—but is **less expected** after *"mike entered the bar [.] he ordered"*

INFORMATION THEORY IN DFS

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

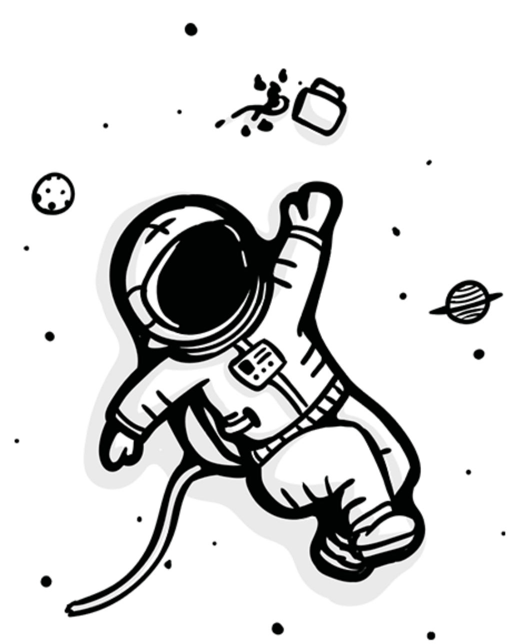
- **Surprisal** quantifies the expectancy of words in context
- Higher Surprisal ⇔ increased processing cost (Hale, 2001; Levy, 2008)
- In DFS, Surprisal quantifies expectancy of transition in meaning space, triggered by message m_{ab}:

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

➡ Word-by-word information effects of semantic constructions

Venhuizen, Crocker, Brouwer (2019, Discourse Proc; 2019, Entropy)

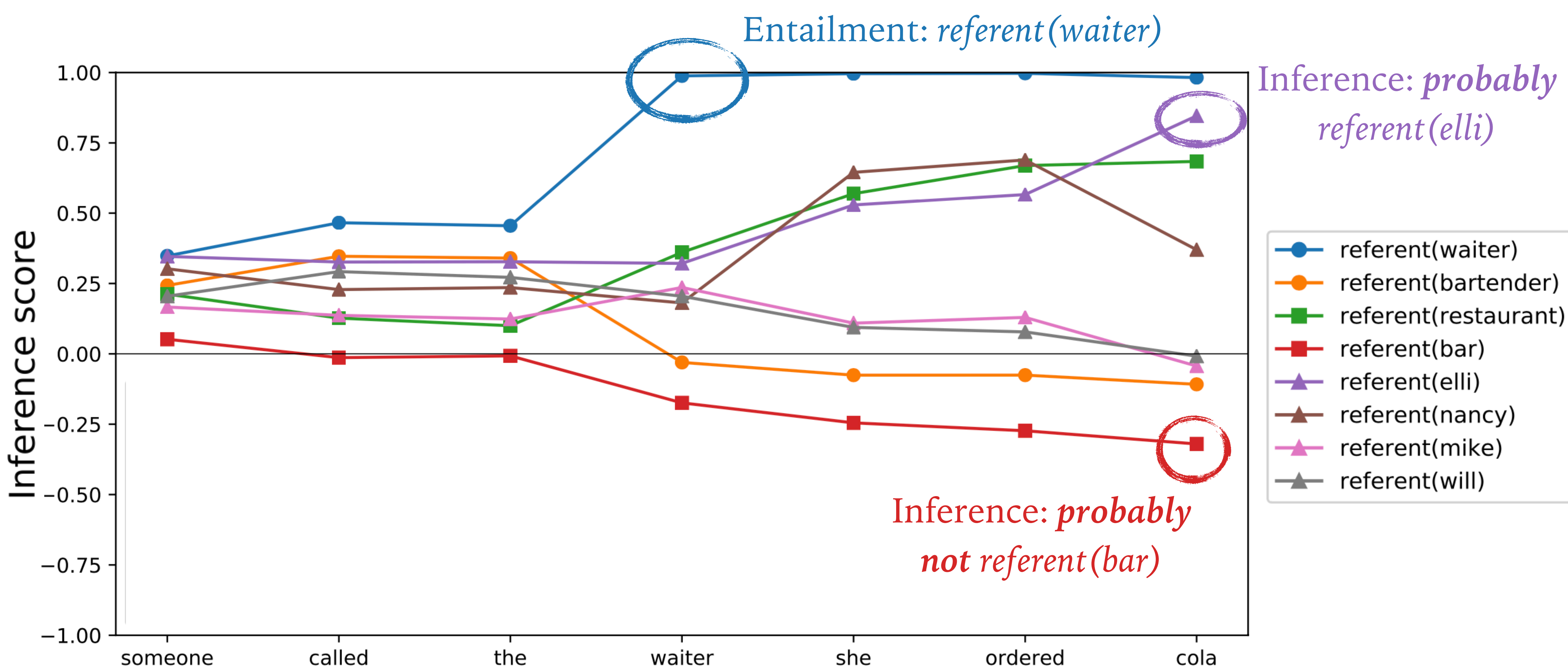
SEMANTIC PROCESSING IN THE MEANING SPACE



ENTAILMENT AND INFERENCE

Incremental meaning construction in the model is driven by:

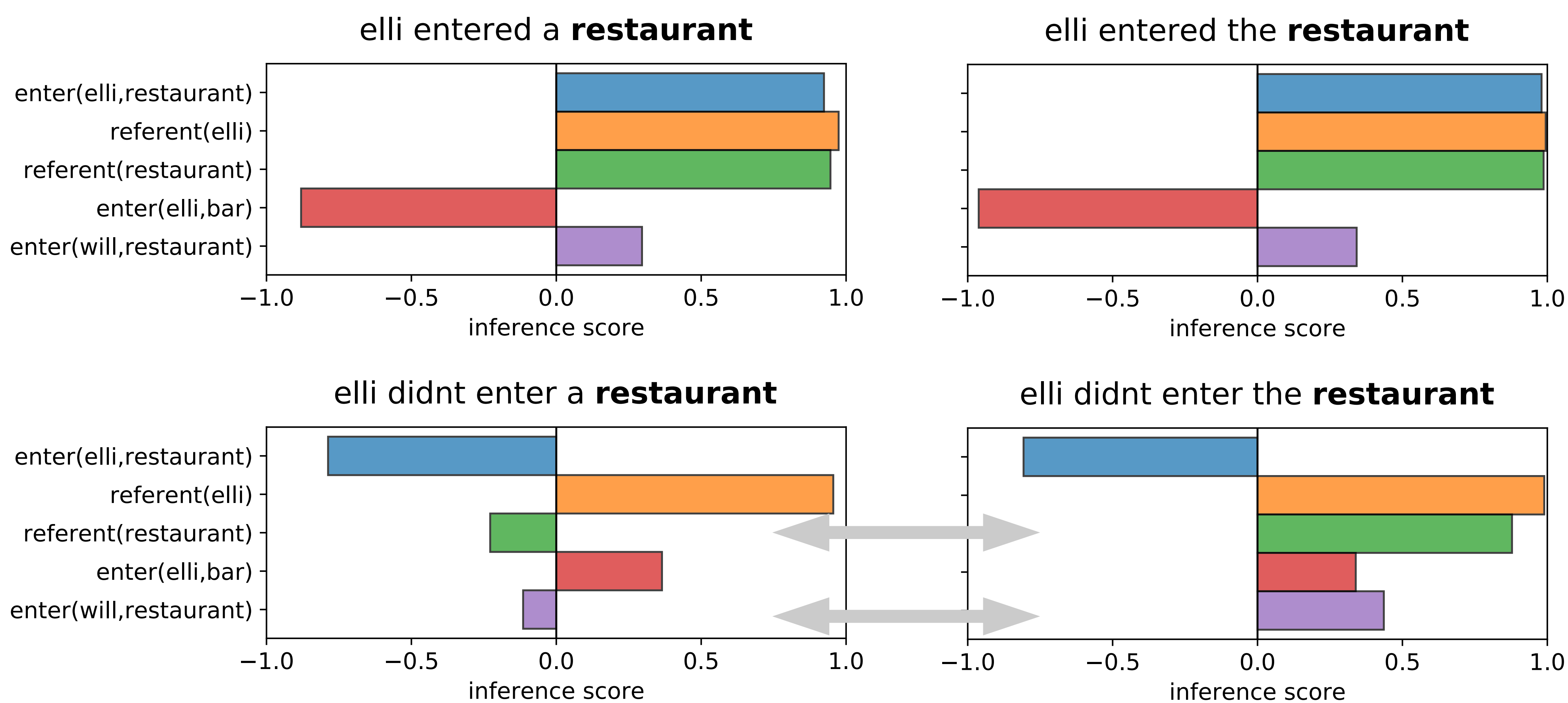
- Sentence-semantics mappings (literal utterance meaning)
- Structure of the meaning space (probabilistic inferences)



Venhuizen, Hendriks, Crocker, Brouwer (2021, Inf. Comp./ArXiv)

NEGATION AND PRESUPPOSITION

- Negation affects entailments and probabilistic inferences
- Interaction between negation and presupposition (triggered by *"the"*)
  - Presupposition has an effect beyond the literal meaning

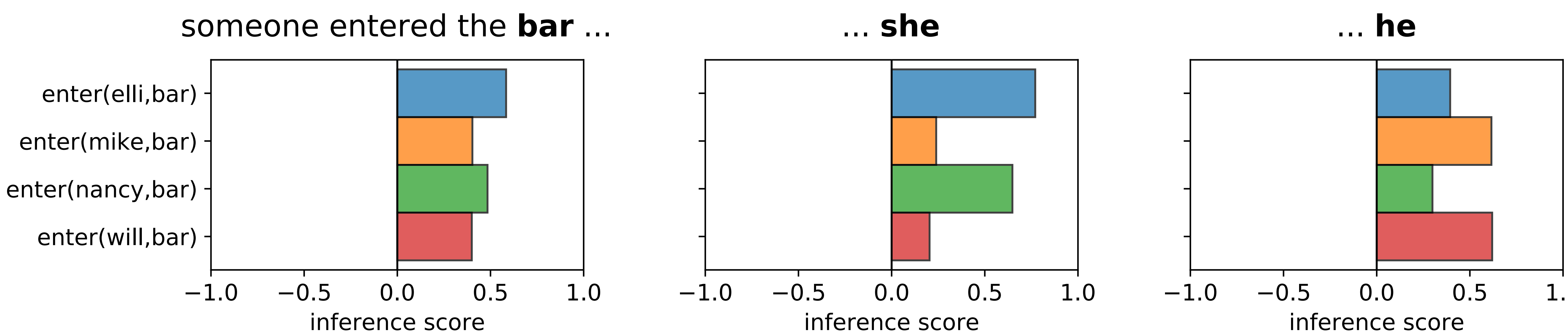


Venhuizen, Hendriks, Crocker, Brouwer (2021, Inf. Comp./ArXiv)

QUANTIFICATION AND REFERENCE

Quantified expressions induce inferential uncertainty

- ➤ Selective expressions (e.g. pronouns) can reduce this uncertainty
- ➤ Confirming initial expectations results in reduced Surprisal



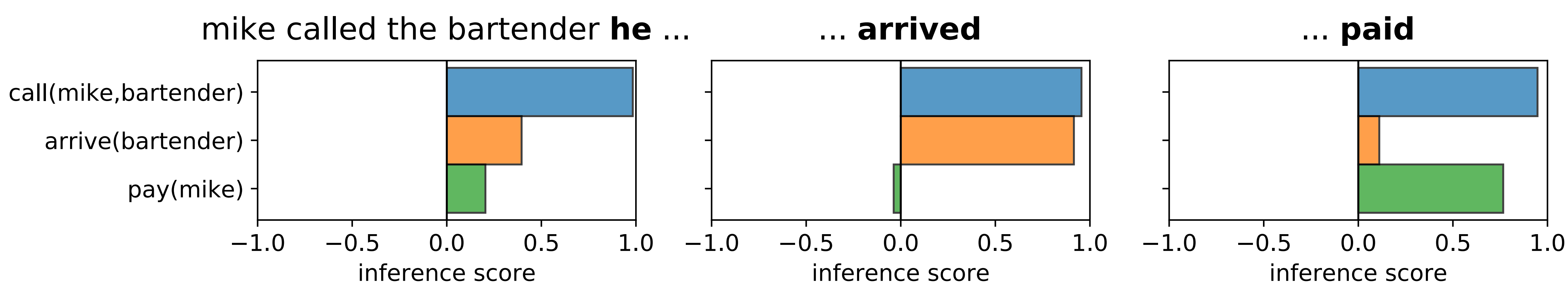
$S(she) = 0.92$ $S(he) = 1.12$

Venhuizen, Hendriks, Crocker, Brouwer (2021, Inf. Comp./ArXiv)

REFERENTIAL AMBIGUITY

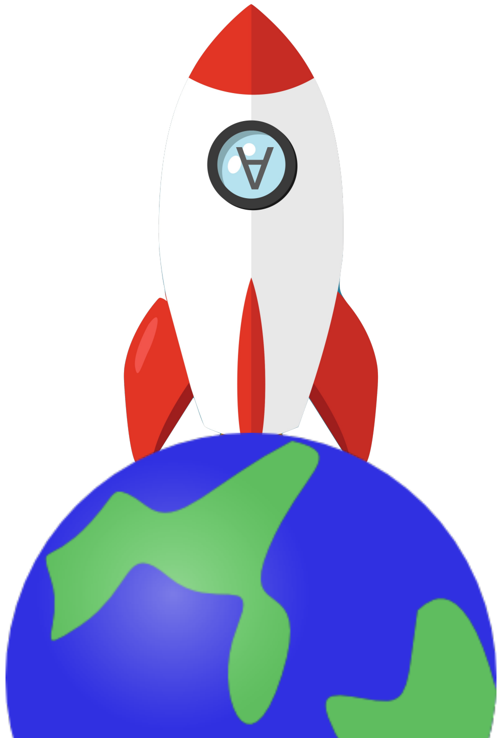
In the training data, the anaphoric antecedent of pronouns is always disambiguated by the preceding or the following context

- Ambiguous pronouns trigger competing hypotheses about the utterance-final interpretation
- Disambiguating continuations result in utterance-level entailments
- Surprisal estimates reflect difference between expected and unexpected continuations



$S(arrived) = 0.44$ $S(paid) = 0.66$

BACK FROM SPACE



SUMMARY

Distributional Formal Semantics

- Compositionality
- Entailment and probabilistic inference
- Incremental meaning construction

?

Distributional Semantics

- Semantic similarity
- Empirically driven
- Cognitively inspired

DS VS. DFS: COMPLEMENTARY ASPECTS OF MEANING

- **Semantic similarity:**

  lexical similarity vs. propositional similarity

  beer ~ wine *order(mike,beer) ~ drink(mike,beer)*

- **Data-driven sampling:**

  bottom-up vs. top-down

  individual linguistic co-occurrences *high-level description of the world*

- **Cognitive foundation:**

  semantic memory vs. utterance interpretation

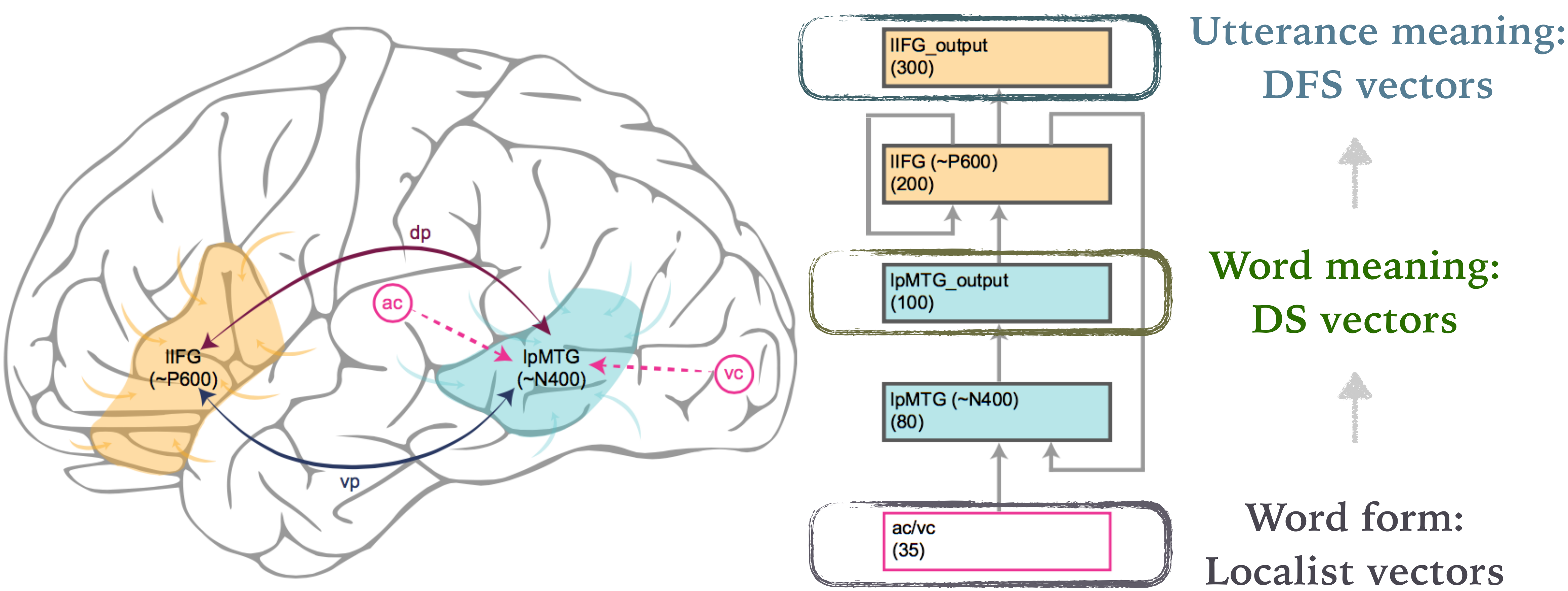
  feature-based word meanings *unfolding discourse-level interpretation*

Kutas & Federmeier (2000); McRae et al. (2005); van Berkum (2009); Brouwer et al. (2012, 2017)

DISCUSSION: COGNITIVE FOUNDATION FOR DFS?

The Retrieval-Integration account of the electrophysiology of language comprehension

- Word meaning retrieval~N400
- Integration in utterance meaning~P600

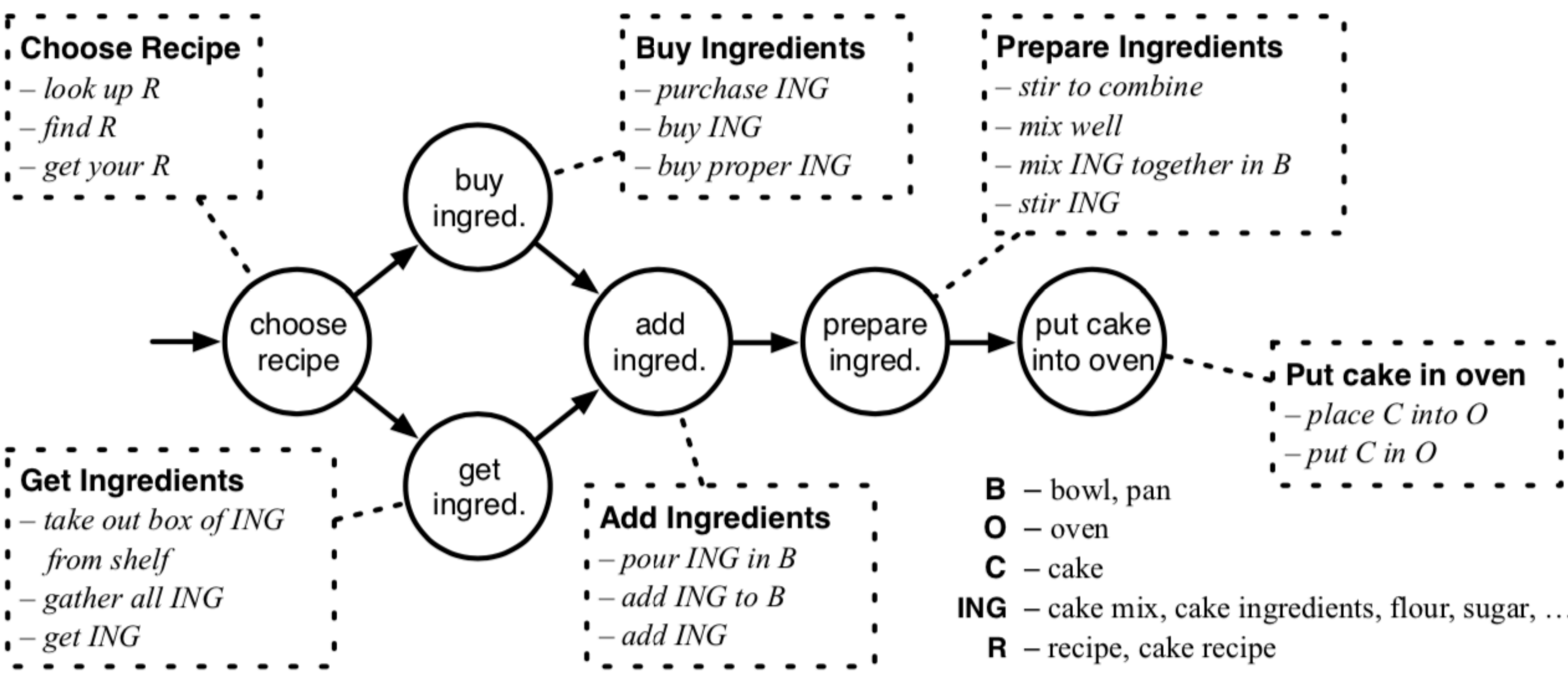


Brouwer et al. (2012, Brain Res.; 2017, Cogn. Sci.; 2021, Frontiers Psych.);

DISCUSSION: DATA-DRIVEN DFS?

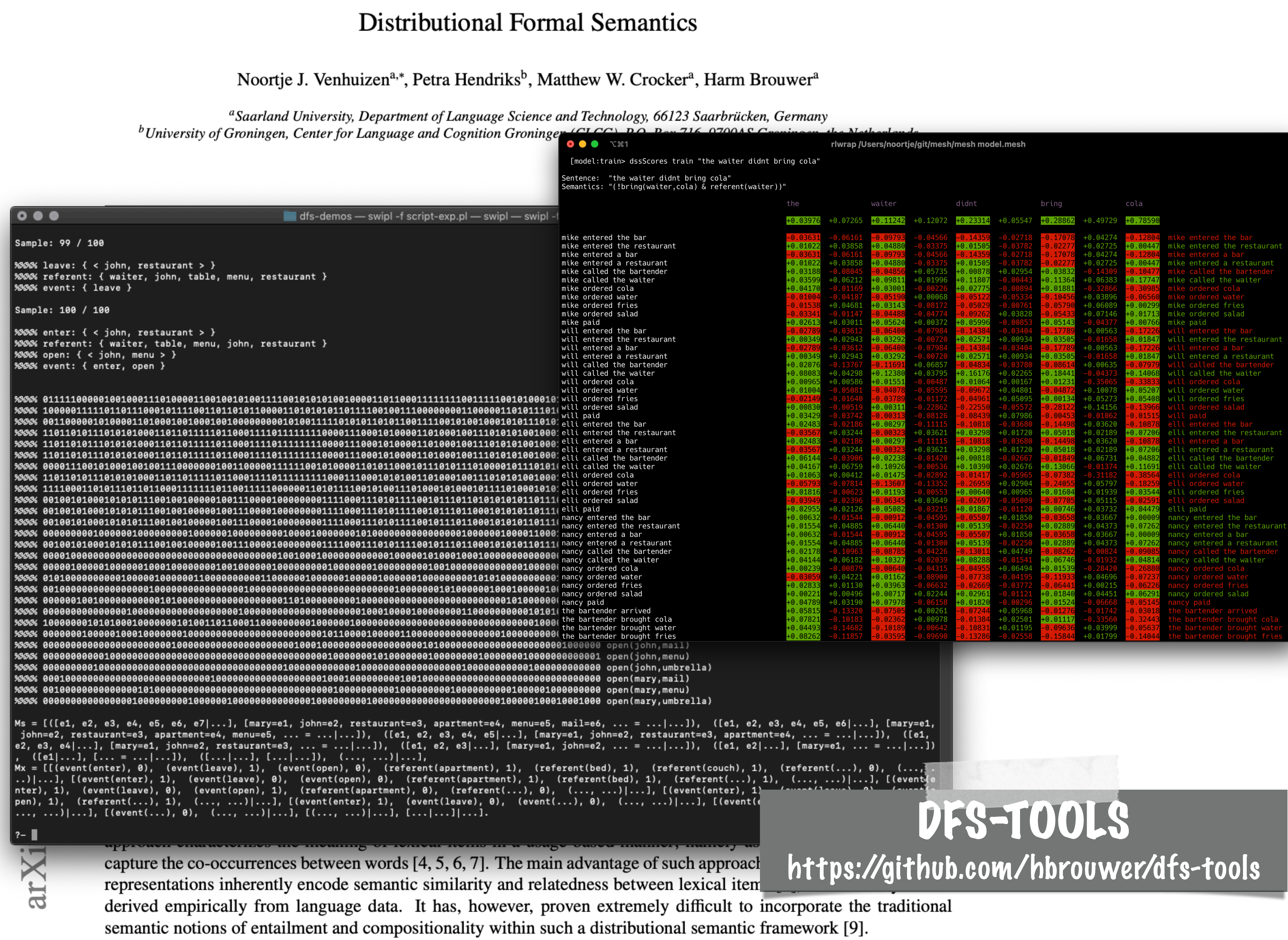
Meaning space reflects world knowledge about propositional co-occurrence, rather than linguistic co-occurrence

- DeScript corpus (Wanzare et al., 2016)



DISTRIBUTIONAL FORMAL SEMANTICS

- The meaning space $S_{\mathcal{M} \times \mathcal{P}}$ captures the structure of the world **truth-conditionally** and **probabilistically**
- Meaning vectors are **compositional** at the propositional level
- **Sub-propositional meaning** derived by incrementally navigating $S_{\mathcal{M} \times \mathcal{P}}$ (using a Simple Recurrent neural Network)
- Semantic phenomena—*negation, presupposition, quantification & reference*—affect **incremental entailments and inferences** during meaning space navigation


Distributional Formal Semantics
Noortje J. Venhuizen^a,*, Petra Hendriks^b, Matthew W. Crocker^a, Harm Brouwer^a
^a Saarland University, Department of Language Science and Technology, 66123 Saarbrücken, Germany
^b University of Groningen, Center for Language and Cognition Groningen
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))"
dfs-demos — swipl -f script-exp.pl — swipl — swipl
Sample: 99 / 100
Sample: 100 / 100
capture the co-occurrences between words [4, 5, 6, 7]. The main advantage of such approach
representations inherently encode semantic similarity and relatedness between lexical item
derived empirically from language data. It has, however, proven extremely difficult to incorporate the traditional
semantic notions of entailment and compositionality within such a distributional semantic framework [9].
DFS-TOOLS
https://github.com/hbrouwer/dfs-tools