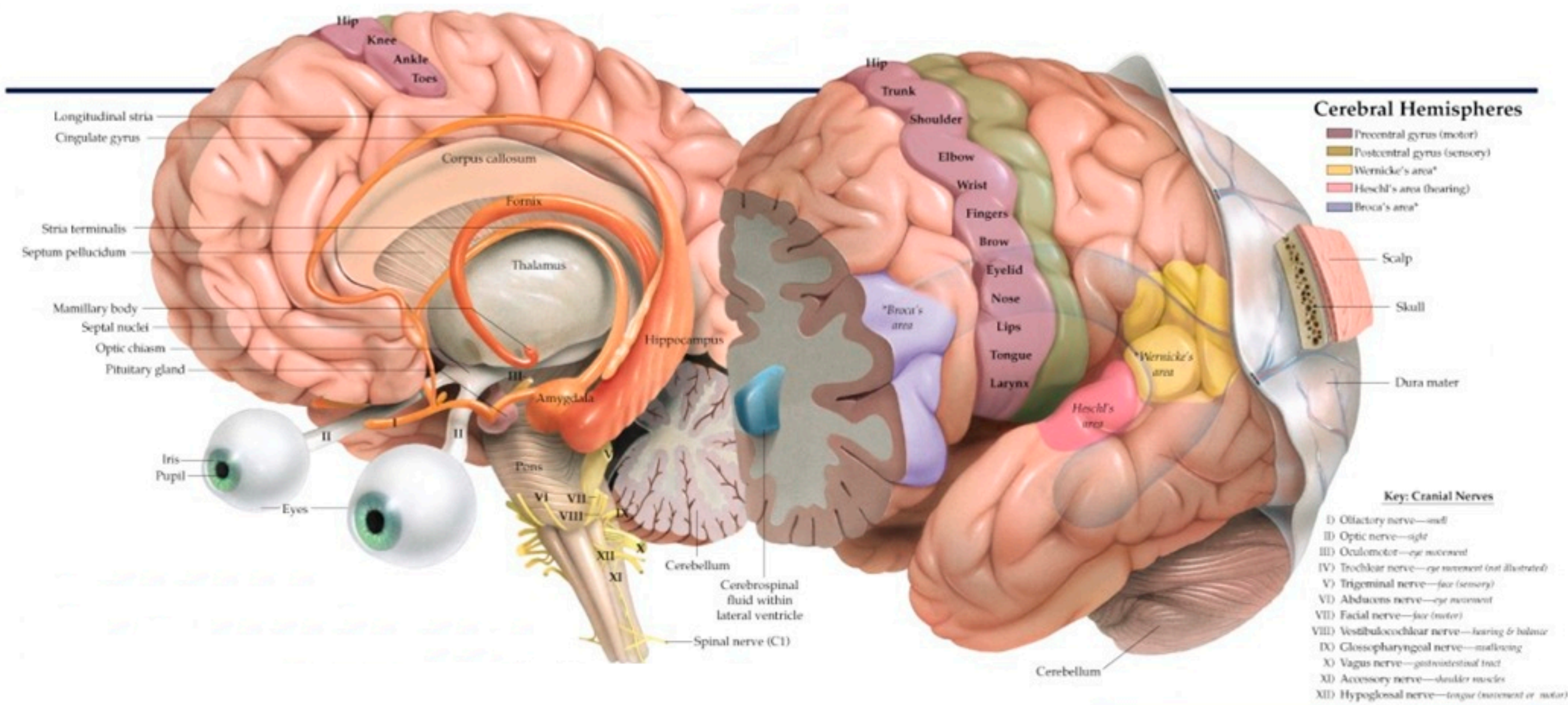
Semantic Theory
Week 10 – Distributional Formal Semantics

Noortje Venhuizen
Harm Brouwer

Universität des Saarlandes

Summer 2021

The Greatest Semanticist of them all …



> Our **language comprehension system** is highly effective and accurate at attributing meaning to unfolding linguistic signal (~word-by-word)

>> This system's **representations** and **computational principles** are implemented in the **neural hardware** of the brain

>>> We should understand meaning construction and representation in terms of "brain-style computation"

A shopping list

Neural Plausibility: assumed representations and computational principles should be implementable at the neural level [cf. Rumelhart, 1989]

Expressivity: representations should capture necessary dimensions of meaning, such as negation, quantification, and modality [cf. Frege, 1892]

Compositionality: the meaning of complex expressions should be derivable from the meaning of its parts [cf. Partee, 1984]

Gradedness: meaning representations are probabilistic, rather than discrete in nature [cf. Spivey, 2008]

Inferential: The derivation of utterance meaning entails (direct) inferences that go beyond literal propositional content [cf. Johnson-Laird, 1983]

Incrementality: As natural language unfolds over time, representations should allow for incremental construction [cf. Tanenhaus et al., 1995]

Brouwer, Crocker & Venhuizen (2017)
In: *From Semantics to Dialectometry*

Distributional Semantics

"How much do we know at any time? Much more, or so I believe, than we know we know!"

— Agatha Christie, *The Moving Finger* (1942)

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

— J. R. Firth (1957)

Psychological Review
1997, Vol. 104, No. 2, 211–240

A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge

Thomas K Landauer
University of Colorado at Boulder

Susan T. Dumais
Bellcore

Distributional Semantics (cont'd)

How much wood would a woodchuck chuck ,
if a woodchuck could chuck wood ?
As much wood as a woodchuck would ,
if a woodchuck could chuck wood .

	a	**as**	**chuck**	**could**	**how**	**if**	**much**	**wood**	**woodch.**	**would**	**,**	**.**	**?**
a	0	5	9	6	1	10	4	8	18	9	10	0	0
as	5	4	2	1	0	0	7	10	3	2	1	0	5
chuck	9	2	0	8	0	5	1	9	11	2	4	3	3
could	6	1	8	0	0	4	0	6	8	0	2	2	2
how	1	0	0	0	0	0	4	3	0	2	0	0	0
if	10	0	5	4	0	0	0	0	10	3	8	0	0
much	4	7	1	0	4	0	0	10	2	3	0	0	3
wood	8	10	9	6	3	0	10	2	8	5	0	4	6
woodch.	18	3	11	8	0	10	2	8	0	8	10	1	1
would	9	2	2	0	2	3	3	5	8	0	5	0	0
,	10	1	4	2	0	8	0	0	10	5	0	0	0
.	0	0	3	2	0	0	0	4	1	0	0	0	0
?	0	5	3	2	0	0	3	6	1	0	0	0	0

(4-word ramped window: 1 2 3 4 [0] 4 3 2 1)

Rohde et al. (under revision) *Cogn. Sci.*

Distributional Semantics (cont'd)

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Ranging from dissimilar (0) to similar (1) — e.g., similarity(wood, woodchuck) = .6

> **Neurally plausible** and **Graded** lexical representations

> But what about **Compositionality**, **Expressivity** and **Inference**?

Queen = King - Man?

X is queen ⊨ X is not a man

X is not a queen = ???

Some queens are rich = ???

→ Distributional Semantics lacks the logical capacity of Formal Semantics

(but is still highly suitable for modelling lexical semantic memory!)

Distributional Formal Semantics

Noortje Venhuizen
Petra Hendriks
Matthew Crocker
Harm Brouwer

NATURAL LANGUAGE SEMANTICS

Model-theoretic Semantics

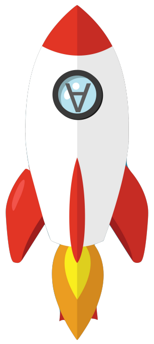
- Truth-conditional meaning
- Logical entailment
- Compositionality

Distributional Semantics

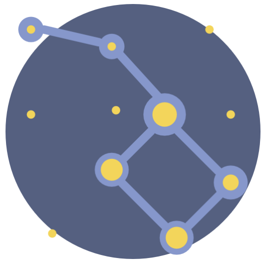
- Semantic similarity
- Empirically driven
- Cognitively inspired

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

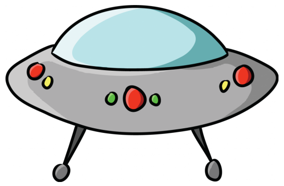
A FRAMEWORK FOR DISTRIBUTIONAL FORMAL SEMANTICS



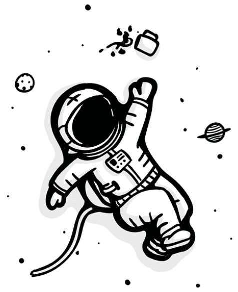
A meaning space for Distributional Formal Semantics



Formal properties of the meaning space

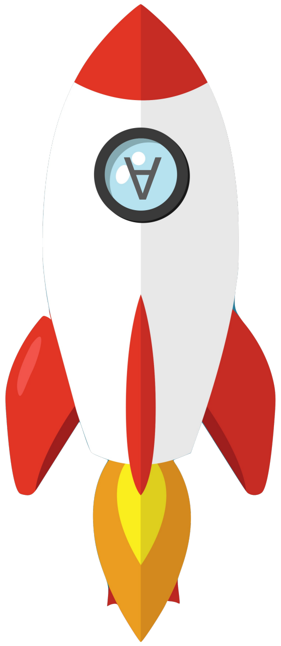


Incremental meaning construction



Semantic processing in the meaning space

A MEANING SPACE FOR DISTRIBUTIONAL FORMAL SEMANTICS



FROM MODELS TO MEANING SPACE

$M_1 = \langle U_1, V_1 \rangle$

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

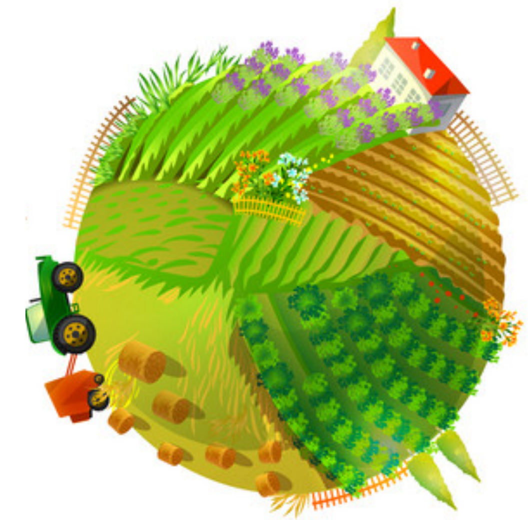
$M_2 = \langle U_2, V_2 \rangle$

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

$M_3 = \langle U_3, V_3 \rangle$

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

…



$M_n = \langle U_n, V_n \rangle$

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

- 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	…
…	…	…	…	…	…

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

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

- **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; 2021, Inf. Comp./ArXiv)

THE DISTRIBUTIONAL HYPOTHESIS REVISITED

"

You shall know a word
by the company it keeps

- J. R. Firth (1957)

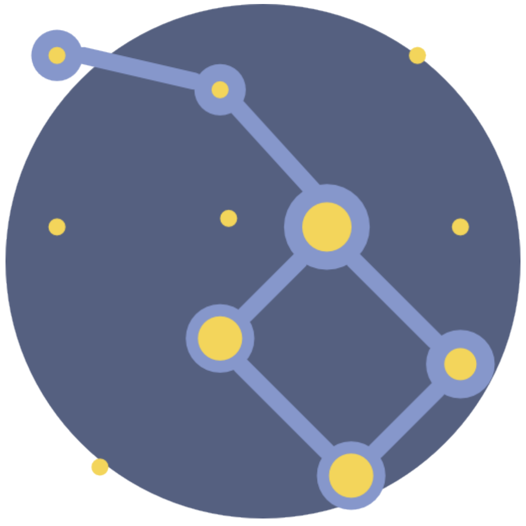
THE DISTRIBUTIONAL HYPOTHESIS REVISITED

“

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

- 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

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

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

… *etc.*

- Quantification is defined relative to the combined universe of $\mathcal{M}_\mathcal{P}$: $\mathcal{U}_\mathcal{M} = \{e_1 \dots e_m\}$ (thereby preserving entailment in $\mathcal{M}_\mathcal{P}$)

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

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

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)

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

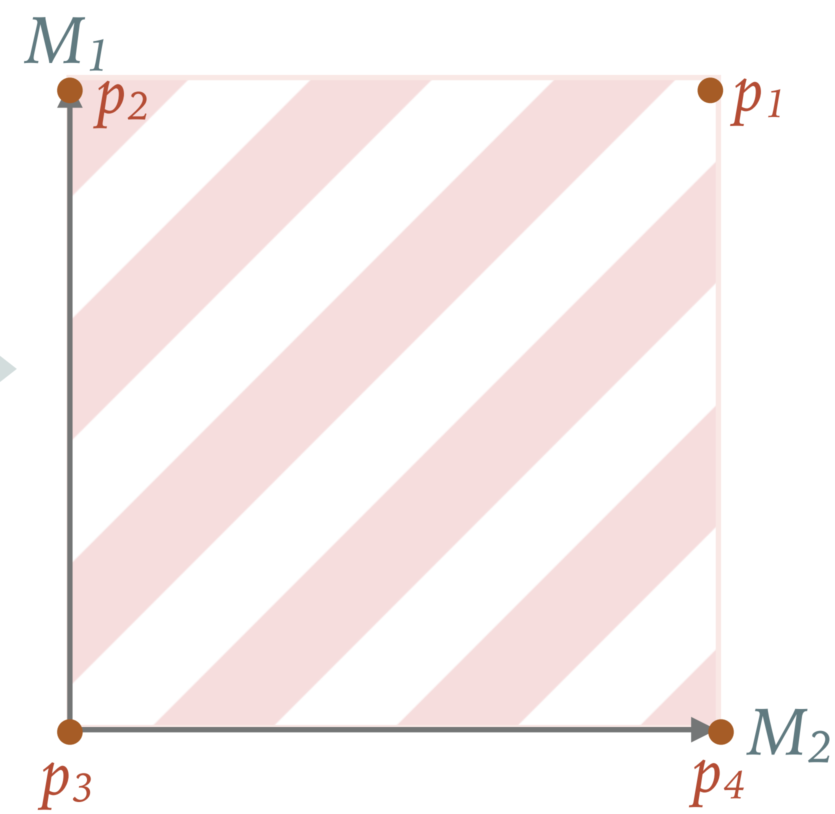
- $P(a|b) \leq P(a)$: Negative inference (b decreases probability of a)

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

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

SAMPLING A MEANING SPACE

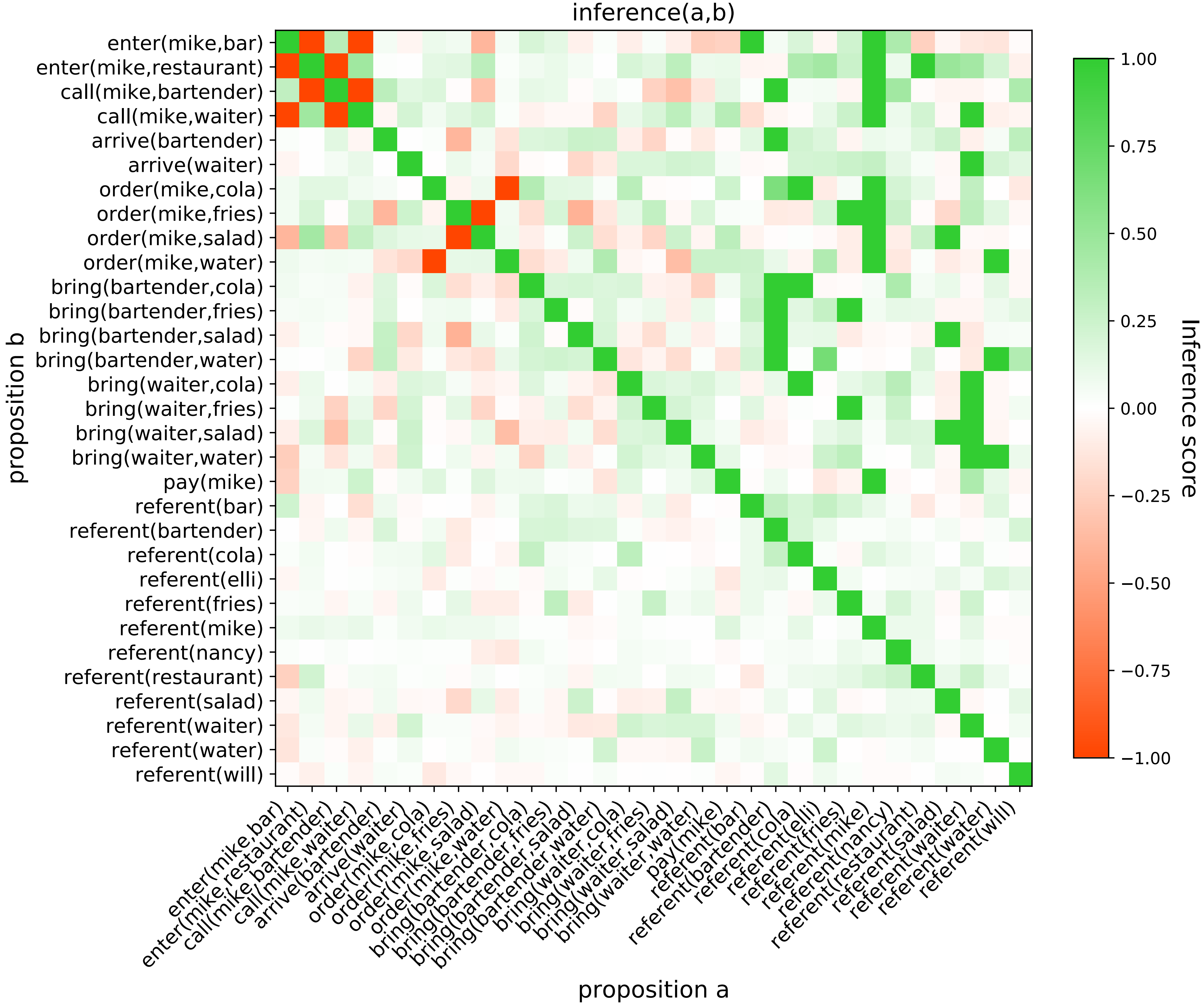
Incremental, inference-based probabilistic sampling: Given a set of propositions $\mathcal{P}$ and a set of constraints C, incrementally construct a model (*Light World*) while keeping track of negative inferences (*Dark World*)

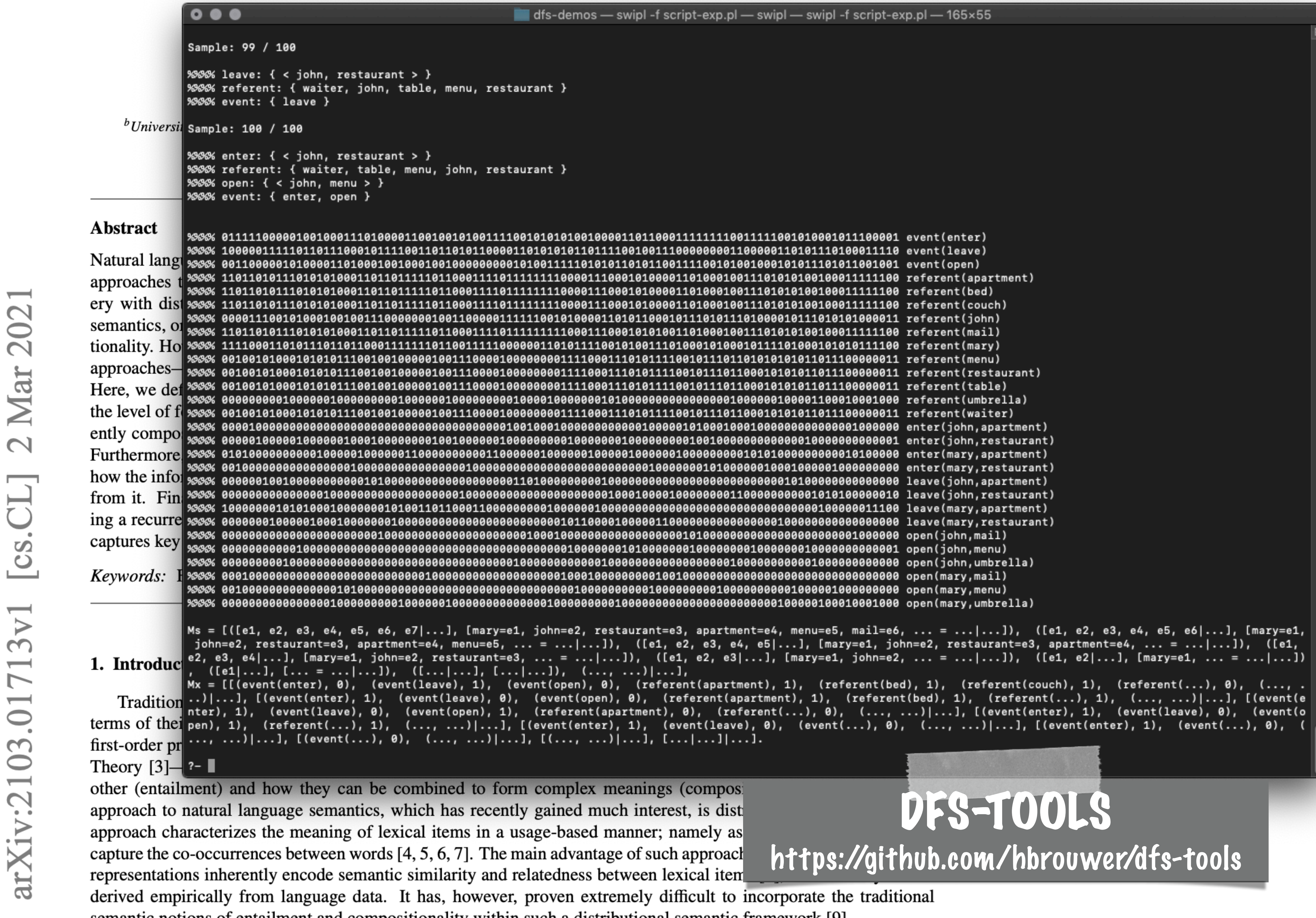
- *a proposition $p \in \mathcal{P}$ is **inferred to be false** iff p can only be true in the Dark World*
  - *p is consistent with respect to the Dark World*
  - *adding p to Light World violates a truth constraint ($c \in C$) on Light World*
- *a proposition $p \in \mathcal{P}$ is **inferred to be true** iff p can only be true in the Light World*
  - *p is consistent with respect to the Light World*
  - *adding p to Dark World satisfies a falsehood constraint ($\bar{c}$, s.t. $c \in C$) on Dark World*
- *if p cannot be inferred, **probabilistic constraints** determine whether p is added to the Light World (with $Pr(p)$) or to the Dark World (with $1-Pr(p)$)*

Venhuizen, Hendriks, Crocker, Brouwer (2021, ArXiv)

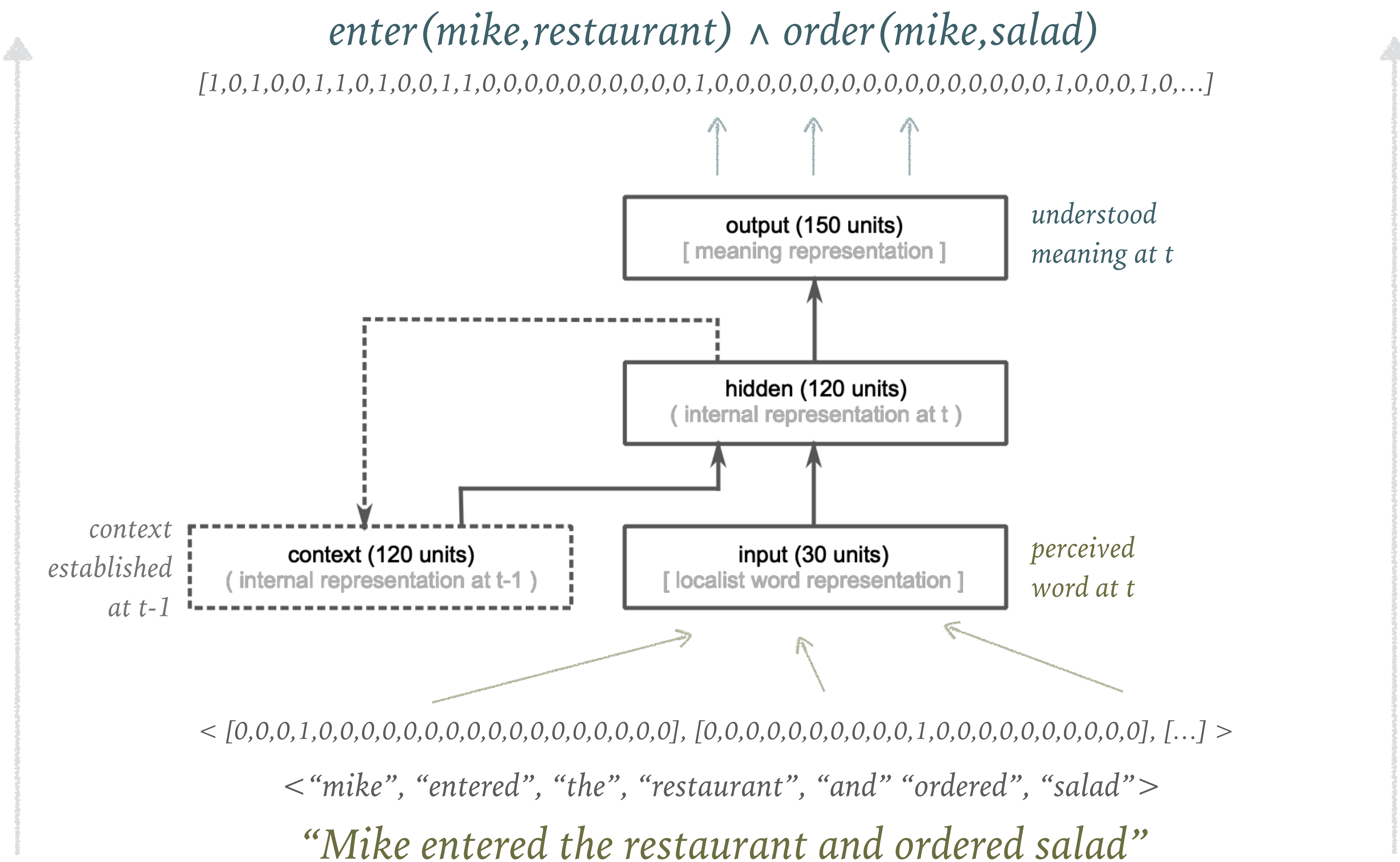
CONSTRUCTING THE MODEL: MEANING SPACE

We sampled a meaning space of 150 models describing 51 propositions



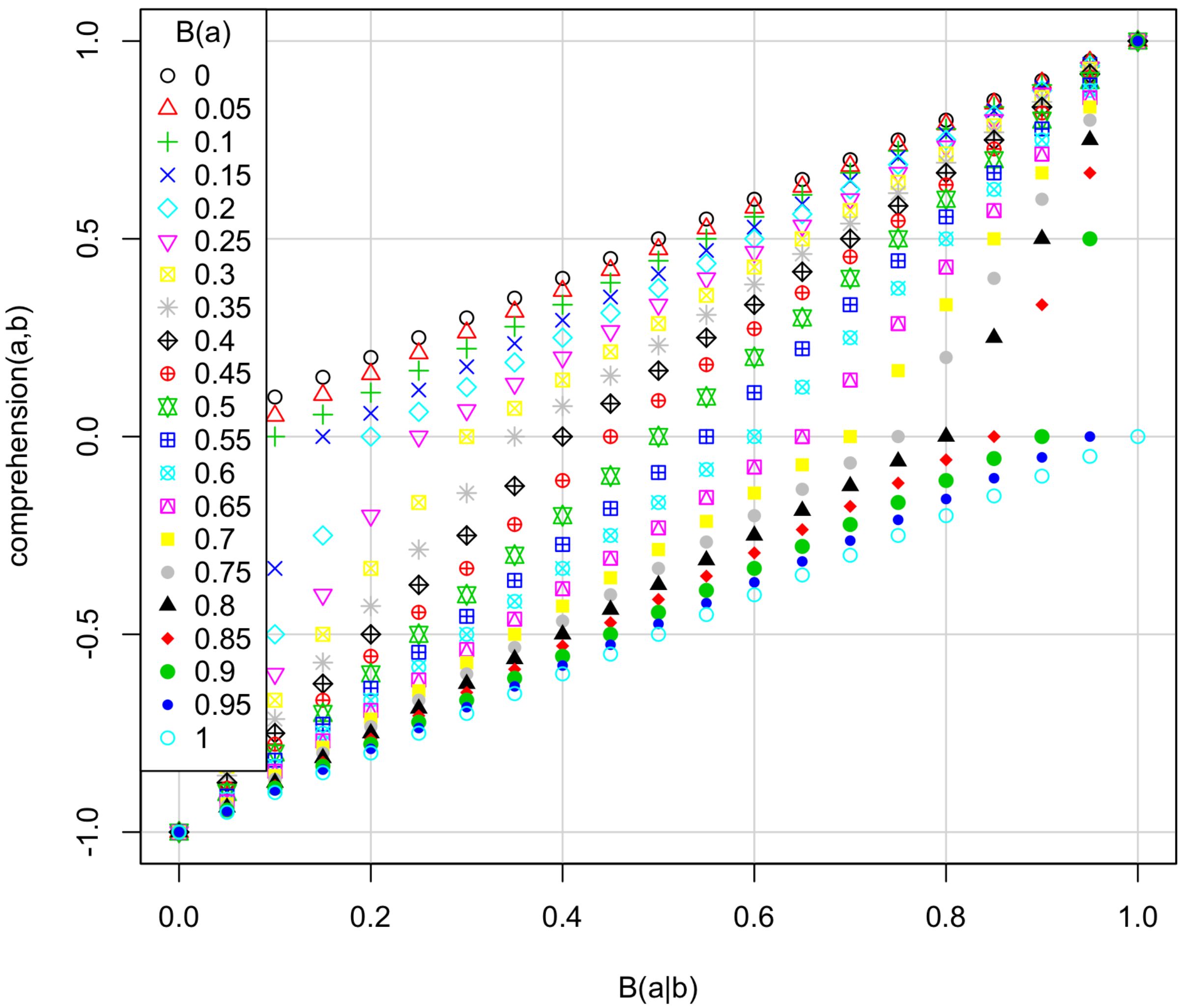

dfs-demos — swipl -f script-exp.pl — swipl — swipl -f script-exp.pl — 165×55
Sample: 99 / 100
%%%% leave: { < john, restaurant > }
%%%% referent: { waiter, john, table, menu, restaurant }
%%%% event: { leave }
Sample: 100 / 100
%%%% enter: { < john, restaurant > }
%%%% referent: { waiter, table, menu, john, restaurant }
%%%% open: { < john, menu > }
%%%% event: { enter, open }
%%%% 0111110000010010001110100001100100101001111001010101001000011011000111111110011111001010001011100001 event(enter)
%%%% 1000001111101101110001011110011011010110000110101010110111100100111000000001100001101011101000111110 event(leave)
%%%% 0011000001010000110100010010001001000000000101001111010101101011001111001010010001010111010110010001 event(open)
%%%% 1101101011101010100011011011111011000111101111111110000111000101000011010001001110101010010001111100 referent(apartment)
%%%% 1101101011101010100011011011111011000111101111111110000111000101000011010001001110101010010001111100 referent(bed)
%%%% 1101101011101010100011011011111011000111101111111110000111000101000011010001001110101010010001111100 referent(couch)
%%%% 0000111001010001001001110000000100110000011111100101000011010110001011101011101000010111010101000011 referent(john)
%%%% 1101101011101010100011011011111011000111101111111110001110001010100110100010011101010100100011111100 referent(mail)
%%%% 1111000110101110110110001111110110011111000000110101111001010011101000101000101111010001010101111100 referent(mary)
%%%% 0010010100010101011100100100000100111000010000000011110001110101111001011101101010101010111000000011 referent(menu)
%%%% 0010010100010101011100100100000100111000010000000011110001110101111001011101100010101011011100000011 referent(restaurant)
%%%% 0010010100010101011100100100000100111000010000000011110001110101111001011101100010101011011100000011 referent(table)
%%%% 0000000001000000100000000010000001000000000100001000000001010000000000000000100000010000110001000001000 referent(umbrella)
%%%% 0010010100010101011100100100000100111000010000000011110001110101111001011101100010101011011100000011 referent(waiter)
%%%% 0000100000000000000000000000000000000100100010000000000000100000101000100010000000000000000001000000 enter(john,apartment)
%%%% 0000010000010000000100010000000001001000000100000000000100000000100000000000100100000000000001000000000001 enter(john,restaurant)
%%%% 0101000000000001000001000000110000000000110000001000000100000100000010000000001010100000000000010100000 enter(mary,apartment)
%%%% 0010000000000000000001000000000000000010000000000000000000000000001000000010100000001000100000001000000000 enter(mary,restaurant)
%%%% 0000001001000000000001010000000000000000000011010000000001000000000000000000000000000001010000000000000000 leave(john,apartment)
%%%% 0000000000000001000000000000000000001000000000000000000000010001000010000000001100000000001010100000010 leave(john,restaurant)
%%%% 1000000010101000100000000101001101100011000000000100000010000000000000000000000000000000000100000011100 leave(mary,apartment)
%%%% 0000001000001000100000000100000000000000000000000010110000100000110000000000000100000000000000000000000 leave(mary,restaurant)
%%%% 0000000000000000000000001000000000000000000000001000100000000000000000001010000000000000000000001000000 open(john,mail)
%%%% 0000000000100000000000000000000000000000000000100000000101000000001000000001000000010000000000000000001 open(john,menu)
%%%% 0000000001000000000000000000000000000001000000000000100000000000000000000100000000000100000000000000000 open(john,umbrella)
%%%% 0001000000000000000000000000000100000000000000000000100010000000000100100000000000000000000000000000000 open(mary,mail)
%%%% 0010000000000000001010000000000000000000000000000000100000000000010000000001000000000100000010000000000 open(mary,menu)
%%%% 0000000000000000001000000000010000000100000000000000000100000000001000000000000000000000000010000100010001000 open(mary,umbrella)
Ms = [([e1, e2, e3, e4, e5, e6, e7|...], [mary=e1, john=e2, restaurant=e3, apartment=e4, menu=e5, mail=e6, ... = ...|...]), ([e1, e2, e3, e4, e5, e6|...], [mary=e1, john=e2, restaurant=e3, apartment=e4, menu=e5, ... = ...|...]), ([e1, e2, e3, e4, e5|...], [mary=e1, john=e2, restaurant=e3, apartment=e4, ... = ...|...]), ([e1, e2, e3, e4|...], [mary=e1, john=e2, restaurant=e3, ... = ...|...]), ([e1, e2, e3|...], [mary=e1, john=e2, ... = ...|...]), ([e1, e2|...], [mary=e1, ... = ...|...]), ([e1|...], [... = ...|...]), ([...|...], [...|...]), (..., ...)|...],
Mx = [[(event(enter), 0), (event(leave), 1), (event(open), 0), (referent(apartment), 1), (referent(bed), 1), (referent(couch), 1), (referent(...), 0), (..., ...)|...], [(event(enter), 1), (event(leave), 0), (event(open), 0), (referent(apartment), 1), (referent(bed), 1), (referent(...), 1), (..., ...)|...], [(event(enter), 1), (event(leave), 0), (event(open), 1), (referent(apartment), 0), (referent(...), 0), (..., ...)|...], [(event(enter), 1), (event(leave), 0), (event(open), 1), (referent(...), 1), (..., ...)|...], [(event(enter), 1), (event(leave), 0), (event(...), 0), (..., ...)|...], [(event(enter), 1), (event(...), 0), (..., ...)|...], [(event(...), 0), (..., ...)|...], [(..., ...)|...], [...|...]|...].
?-
DFS-TOOLS
https://github.com/hbrouwer/dfs-tools
arXiv:2103.01713v1 [cs.CL] 2 Mar 2021
[b]Universi
Abstract
Natural lang
approaches t
ery with dist
semantics, or
tionality. Ho
approaches—
Here, we def
the level of f
ently compo
Furthermore
how the infor
from it. Fin
ing a recurre
captures key
Keywords: I
1. Introduc
Tradition
terms of thei
first-order pr
Theory [3]—
other (entailment) and how they can be combined to form complex meanings (compos
approach to natural language semantics, which has recently gained much interest, is dist
approach characterizes the meaning of lexical items in a usage-based manner; namely as
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].

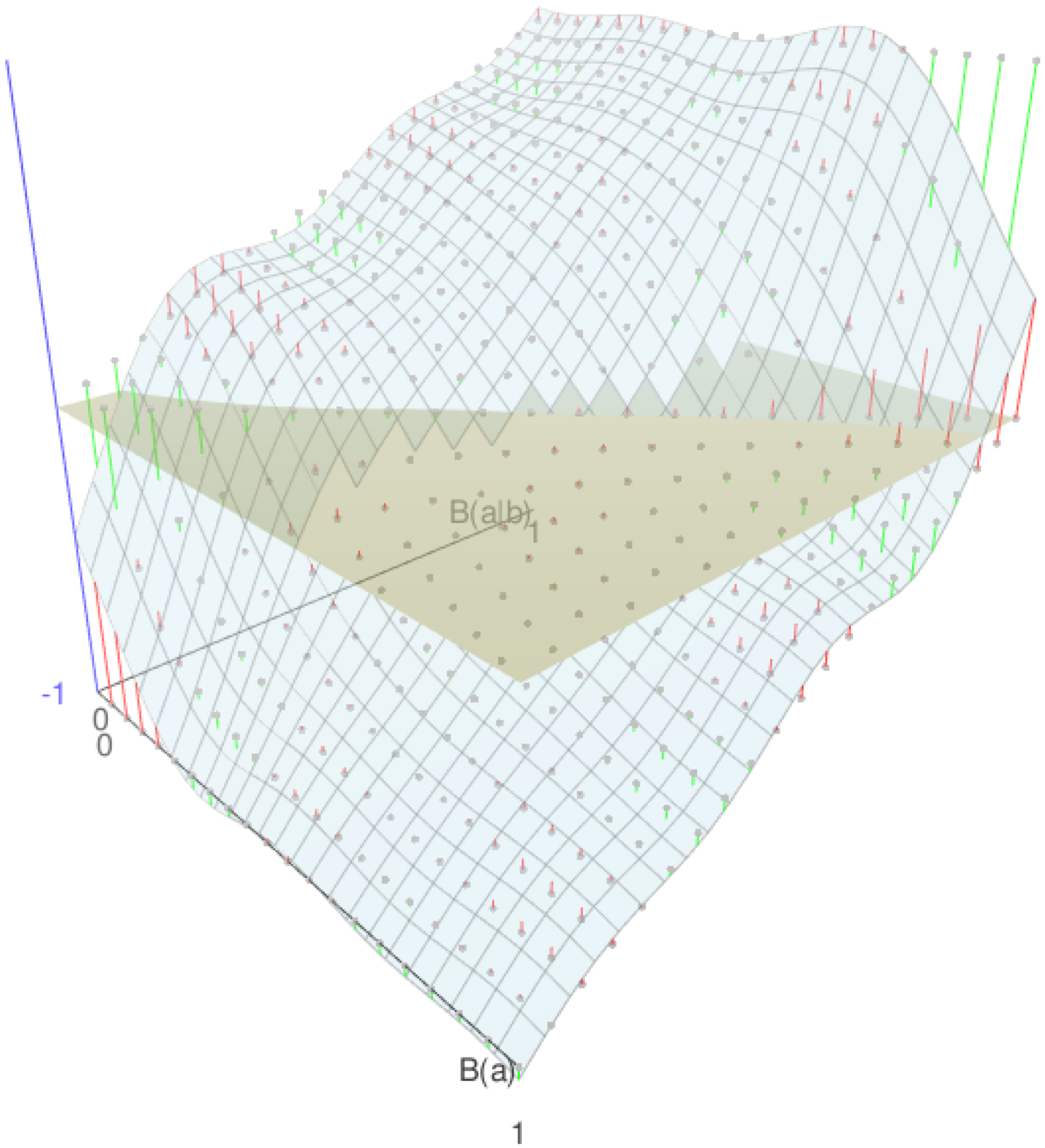
A MODEL OF INCREMENTAL MEANING CONSTRUCTION



— to be continued —

Comprehension scores





The higher *B(a)* the more difficult it is to *increase certainty* in *a*, and the lower *B(a)* the more difficult it is to *increase uncertainty* in *a*