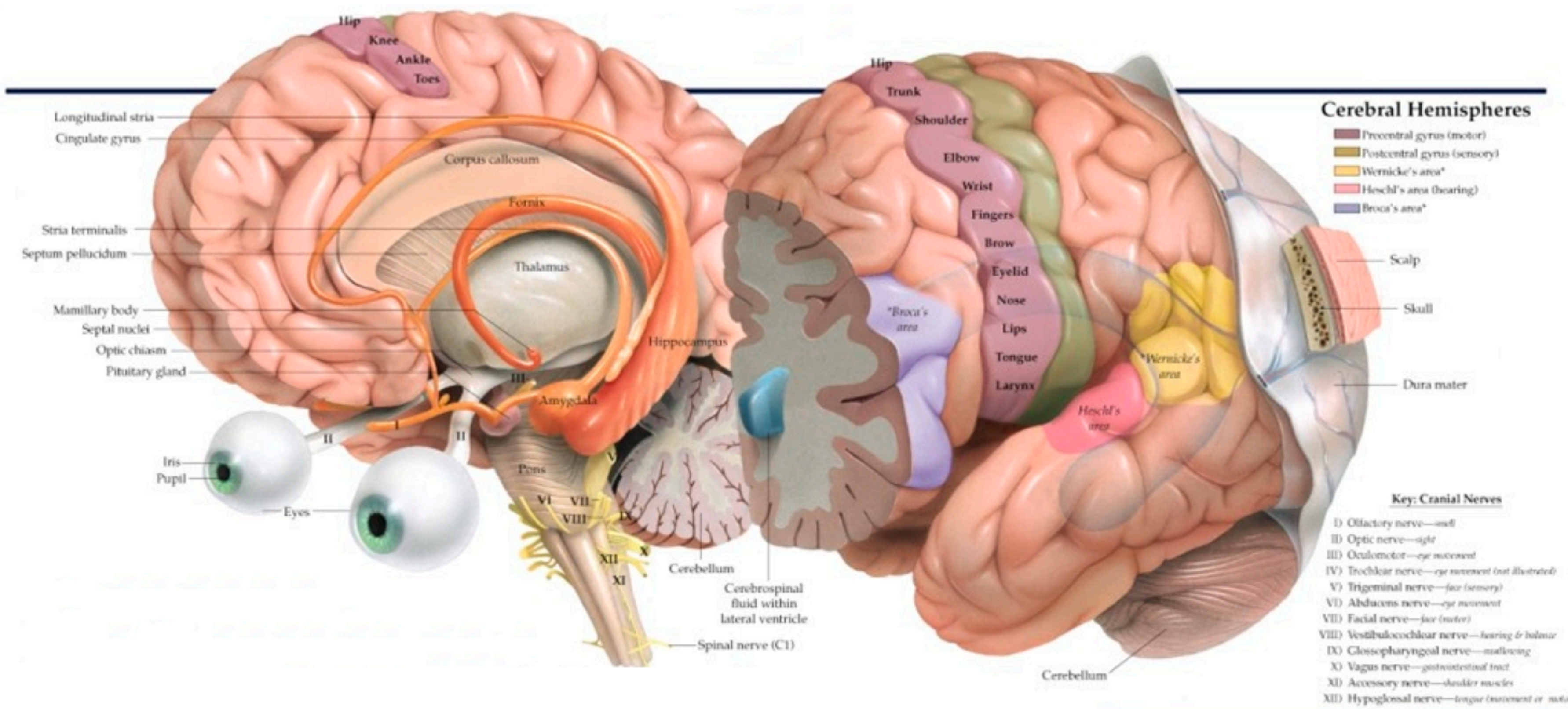
# Semantic Theory
# Week 10 – Distributional Formal Semantics

Noortje Venhuizen
Harm Brouwer

Universität des Saarlandes

Summer 2019

# 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


0033-295X/97/$3.00

**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 not a queen = ???

X is queen ⊨ X is not a man

Some queens are rich = ???

→ Distributional Semantics lacks the logical capacity of Formal Semantics

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

# A FRAMEWORK FOR DISTRIBUTIONAL FORMAL SEMANTICS

*Noortje Venhuizen*
*Petra Hendriks*
*Matthew Crocker*
*Harm Brouwer*

**Model-theoretic Semantics**

- Truth-conditional meaning
- Logical entailment
- Compositionality

?

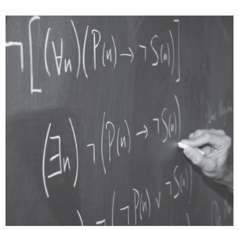
**Distributional Semantics**

- Semantic similarity
- Empirically driven
- Cognitively inspired

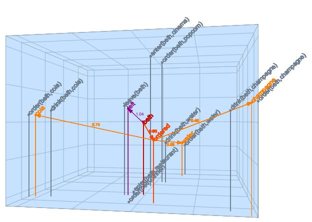
# A FRAMEWORK FOR DISTRIBUTIONAL FORMAL SEMANTICS

A formal distributional meaning space

Sampling a meaning space



Formal properties of the meaning space



Incremental meaning construction (**next week!**)

"

You shall know a word
by the company it keeps

*- J. R. Firth (1957)*

"

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

*- J. R. Firth (1957)*

# FROM MODELS TO MEANING SPACE

$M_1 = \langle U_1, V_1 \rangle$

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

$M_2 = \langle U_2, V_2 \rangle$

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

$M_3 = \langle U_3, V_3 \rangle$

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

…

$M_n = \langle U_n, V_n \rangle$

$p_1 \wedge p_2 \wedge \ldots$

- Together, the set of models $\mathcal{M}$ and the set of propositions $\mathcal{P}$ define the **meaning space** $S_{\mathcal{M}\times\mathcal{P}}$
- Propositional meaning defined by **co-occurrence** across models

# CAPTURING THE STRUCTURE OF THE WORLD

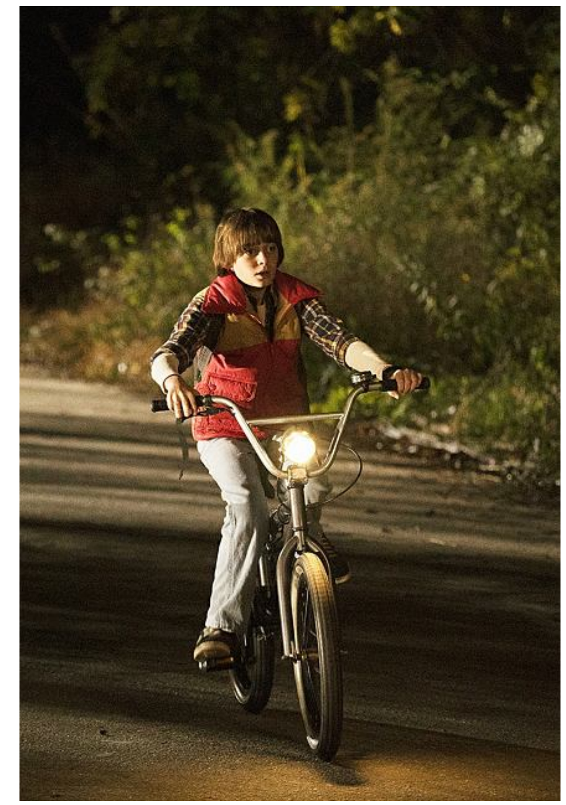
**Will rides a bike**

*The bike has two wheels*

*Will is (likely) outside*

*If its dark, his light is on*

…



- Propositional co-occurrence in $S_{\mathcal{M}\times\mathcal{P}}$ is constrained by a set of **hard** and **probabilistic** world knowledge constraints
  - $p_i$ and $p_j$ never/always co-occur (hard constraint)
  - if $p_k$ holds, then $p_m$ is more likely than than $p_n$ (probabilistic constraint)

# SAMPLING A MEANING SPACE

**Incremental, inference-based probabilistic sampling:**

Given a set of propositions $\mathcal{P}$, construct a model (*Light World*) while keeping track of false propositions (*Dark World*)

- *proposition p is **inferred to be false** iff p can only be true wrt the Dark World*
    - *p is consistent with respect to the dark world*
    - *adding p to Light World violates truth-constraints on Light World*
- *proposition p is **inferred to be true** iff p can only be true wrt the Light World*
    - *p is consistent with respect to the light world*
    - *adding p to Dark World satisfies falsehood-constraints on Dark World*
- *if p cannot be inferred, its truth value is determined **probabilistically***

The Legend of Zelda: A Link to the Past (Nintendo, 1992)

# SAMPLING A MEANING SPACE: EXAMPLE

**Truth-constraint:** LW ⊨ *All boys ride a bicycle*

*? Mike rides a bicycle*

*Dustin rides a bicycle*

*Lucas rides a bicycle*

*Mike rides a bicycle*

*Mike rides a bicycle*

**Falsehood-constraint:** DW ⊨ *There is a boy that rides a bicycle*

# SAMPLING A MEANING SPACE: EXAMPLE

**Truth-constraint:** LW ⊨ *All boys ride a bicycle*

✓ *Mike rides a bicycle*

*Dustin rides a bicycle*

*Lucas rides a bicycle*

*Mike rides a bicycle*

**Falsehood-constraint:** DW ⊨ *There is a boy that rides a bicycle*

# DFS MEANING SPACE

For a set of models $\mathcal{M}$ and a set of propositions $\mathcal{P}$, define the 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 | … |
| | 0 | 1 | 0 | 0 | … |
| | … | … | … | … | … |

*meaning of* $p_n$: $v(p_n)$

*where* $v_i(p_n) = 1$ *iff* $M_i \vDash p$

- *Co-occurrence defines meaning:* Propositions with related meanings will be true in many of the same models

# FORMAL PROPERTIES OF $S_{\mathcal{M}\times\mathcal{P}}$ — COMPOSITIONALITY

Meaning vectors can be combined compositionally

- Negation and conjunction provide functional completeness

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

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

- Quantification is defined over the combined universe of $\mathcal{M}$: $U_{\mathcal{M}}=\{u_1 \dots u_k\}$

  $v_i(\forall x\varphi) = 1 \text{ iff } M_i \vDash \varphi[x\backslash u_1] \wedge \dots \wedge \varphi[x\backslash u_k]$

  $v_i(\exists x\varphi) = 1 \text{ iff } M_i \vDash \varphi[x\backslash u_1] \vee \dots \vee \varphi[x\backslash u_k]$

# FORMAL PROPERTIES OF $S_{\mathcal{M}\times\mathcal{P}}$ — PROBABILITY

Meaning vectors inherently encode (co-)occurrence probabilities

- Prior probability of proposition $p$

  $$P(p) = |\{M_i \in \mathcal{M} \mid M_i \vDash p\}| / |\mathcal{M}|$$

- Given the compositional nature of $S_{\mathcal{M}\times\mathcal{P}}$, the (prior) probability of any formula $\varphi$ can be defined, e.g.:

  $$P(p \wedge q) = |\{M_i \in \mathcal{M} \mid M_i \vDash p \;\&\; M_i \vDash q\}| / |\mathcal{M}|$$

- Conditional probability of formula $\psi$ given $\varphi$

  $$P(\psi | \varphi) = P(\varphi \wedge \psi) \,/\, P(\varphi)$$

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ | |
|---|---|---|---|---|---|
| $M_1$ | 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 | … |
| | … | … | … | … | … |

# FORMAL PROPERTIES OF $S_{\mathcal{M}\times\mathcal{P}}$ — INFERENCE

Probabilistic logical inference of formula $\psi$ given $\varphi$

$$inf(\psi,\varphi) = \begin{cases} [P(\psi \mid \varphi) - P(\psi)] \,/\, [1 - P(\psi)] & \text{if } P(\psi \mid \varphi) > P(\psi) \\ [P(\psi \mid \varphi) - P(\psi)] \,/\, P(\psi) & \text{otherwise} \end{cases}$$

- $P(\psi \mid \varphi) > P(\psi)$: Positive inference ($\varphi$ increases probability of $\psi$)

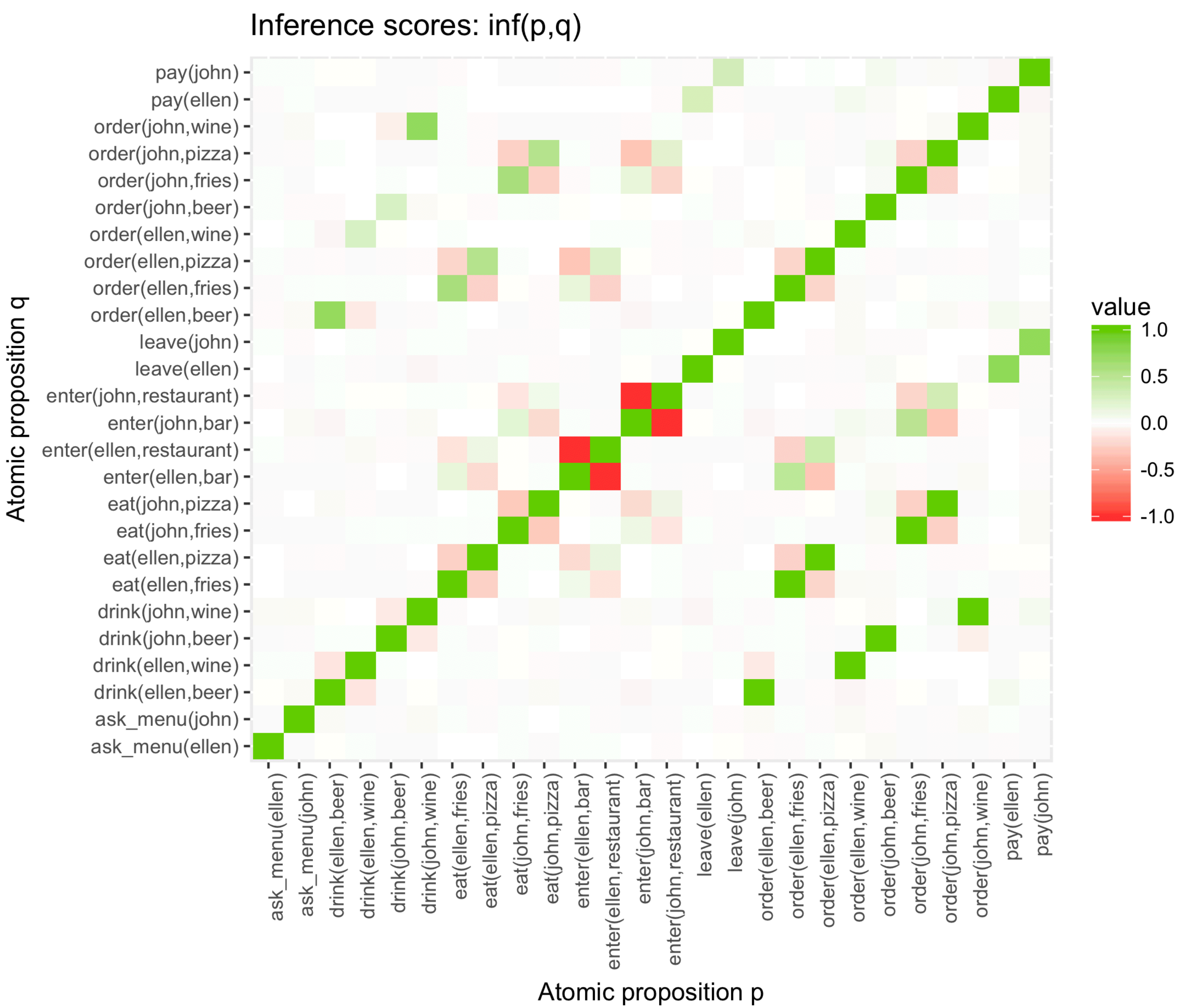
$$inf(\psi,\varphi) = 1 \Leftrightarrow \varphi \vDash \psi$$

- $P(\psi \mid \varphi) \leq P(\psi)$: Negative inference ($\varphi$ decreases probability of $\psi$)

$$inf(\psi,\varphi) = -1 \Leftrightarrow \varphi \vDash \neg\psi$$

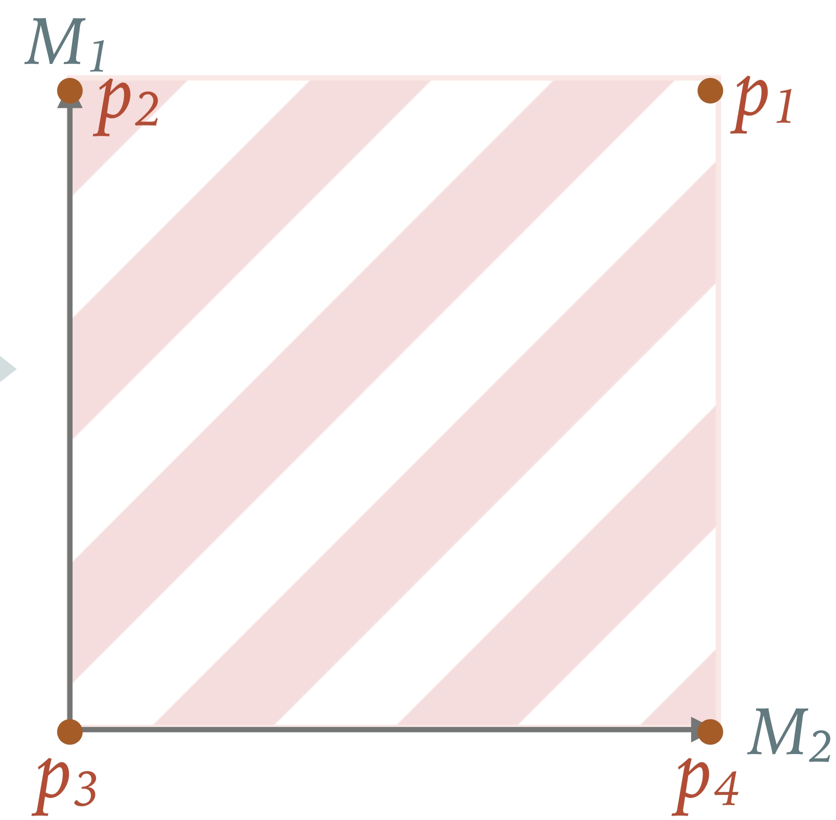
Frank et al. (2003)
*Cognitive Sci.*

# WORLD KNOWLEDGE INFERENCING IN $S_{\mathcal{M}\times\mathcal{P}}$



# SUB-PROPOSITIONAL MEANING IN $S_{\mathcal{M}\times\mathcal{P}}$

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ | |
|---|---|---|---|---|---|
| $M_1$ | 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 | … |
| | … | … | … | … | … |




- Continuous nature of $S_{\mathcal{M}\times\mathcal{P}}$ allows for defining sub-propositional meaning
- Sub-propositional meaning derives from incremental mapping from (sequences of) words to propositions
- **More on this next week!**

**Distributional Formal Semantics**

- Compositionality
- Probabilistic inference
- Incremental meaning construction

?

**Distributional Semantics**

- Semantic similarity
- Empirically driven
- Cognitively inspired

# DISTRIBUTIONAL VS. DISTRIBUTIONAL FORMAL SEMANTICS

- **Semantic similarity**: lexical vs. propositional

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

- **Data-driven sampling**: bottom-up vs. top-down

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

- **Neurocognition**: lexical retrieval vs. semantic integration

  *feature-based word meanings* *unfolding utterance interpretation*

# DISTRIBUTIONAL FORMAL SEMANTICS

- 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 constructed by **incrementally** navigating $S_{\mathcal{M}\times\mathcal{P}}$ (e.g., using an SRN)
- Meaning space navigation reflects direct **pragmatic inference**

DFS-TOOLS
https://github.com/hbrouwer/dfs-tools
ask for a demo!

# MODELING SEMANTIC AND PRAGMATIC THEORY (I)

- **Reference** [Van Berkum et al., 2004]

  1-ref: David shot at Linda as <u>he</u> jumped … ↦ shot(d,l) & jump(d)

  0-ref: Anna shot at Linda as <u>he</u> jumped … ↦ shot(a,l) & (jump(d) | jump(j) | …)

  2-ref: David shot at John as <u>he</u> jumped … ↦ shot(d,j) & ???? [~Nref]

- **Inference — bridging / presupposition** [Burkhardt, 2007]

| | | | |
|---|---|---|---|
| Student <u>shot</u> | + <u>the pistol</u> | ↦ | inf(instr(pistol),shot) ~ +1 |
| Student <u>killed</u> | + <u>the pistol</u> | ↦ | 0 < inf(instr(pistol),killed) < +1 |
| Student <u>found dead</u> | + <u>the pistol</u> | ↦ | 0 < inf(instr(pistol),found_dead) < +1 |
| No student <u>shot</u> | + <u>the pistol</u> | ↦ | inf(instr(pistol),!shot) ~ −1 |

- **Quantification** [e.g., Spychalska et al., 2015]

| | | |
|---|---|---|
| Someone entered restaurant | ↦ | enter(a,r) \| enter(b,r) \| enter(c,r) |
| Two people entered restaurant | ↦ | (enter(a,r) & enter(b,r)) \| (enter(b,r) & …) \| … |
| All people entered restaurant | ↦ | enter(a,r) & enter(b,r) & enter(c,r) |

# MODELING SEMANTIC AND PRAGMATIC THEORY (II)

- Inference — non-literal meaning (irony/metaphor) [Regel et al., 2011]

| | | | |
|---|---|---|---|
| [no context] | + These artists are gifted | ↦ | are_gifted(a) |
| Good performance | + These artists are gifted | ↦ | … & are_gifted(a) |
| Bad performance | + These artists are gifted | ↦ | … & suck_big_time(a) |

- Scalar implicatures [Noveck & Posada, 2003; Nieuwland et al., 2010; Hunt III et al., 2012]

| | | | |
|---|---|---|---|
| Context: | All Xs are P | ↦ | p(x1) & p(x2) & p(x3) |
| Logical: | Some Xs are P | ↦ | p(x1) \| … \| (p(x2) & p(x3)) \| (p(x1) & p(x2) & p(x3)) |
| Pragmatic: | Some Xs are P | ↦ | (p(x1) \| … \| (p(x2) & p(x3))) & !(p(x1) & p(x2) & p(x3)) |

- Negation [Fischler et al., 1983]

| | | | |
|---|---|---|---|
| True affirmative: | A robin is a bird | ↦ | inf(bird,robin) ~ +1 |
| False affirmative: | A robin is a vehicle | ↦ | inf(vehicle,robin) ~ −1 |
| True negative: | A robin is not a vehicle | ↦ | inf(!vehicle,robin) ~ +1 |
| False negative: | A robin is not a bird | ↦ | inf(!bird,robin) ~ −1 |

# SAMPLING A MEANING SPACE

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

*Based on a set of propositions $P$ and a set of constraints $C$, for any proposition $p \in P$*

- *Light World consistent* $(LWC = \top)$ *iff* $\forall c \in C$: $LW_p \vDash c$ *or* $DW \nvDash \bar{c}$
- *Dark World consistent* $(DWC = \top)$ *iff* $\forall c \in C$: $LW \vDash c$ *or* $DW_p \nvDash \bar{c}$
    - $LWC$ & $DWC$ $\Rightarrow$ *p inferred with probability* $Pr(p)$
    - $LWC$ & $\neg DWC$ $\Rightarrow$ *p inferred to Light World*
    - $\neg LWC$ & $DWC$ $\Rightarrow$ *p inferred to Dark World*
    - $\neg LWC$ & $\neg DWC$ $\Rightarrow$ *inconsistent model*

# SAMPLING A MEANING SPACE: INCONSISTENCY EXAMPLE

**Truth-constraints:** LW ⊨ *A boy rides a bicycle*

LW ⊨ *Mike does not ride a bicycle*

*? Mike rides a bicycle*

*Dustin rides a bicycle*

*Lucas rides a bicycle*

**Falsehood-constraints:** DW ⊨ *All boys ride a bicycle*

DW ⊨ *Mike does not ride a bicycle*

# FALSEHOOD CONSTRAINTS (COMPLEMENT)

The complement of any well-formed formula is recursively defined as follows (where ϕ’ is the complement of ϕ):

$p \mapsto p$

$\neg\phi \mapsto \neg\phi'$

$\phi\wedge\psi \mapsto \phi'\vee\psi'$

$\phi\vee\psi \mapsto \phi'\wedge\psi'$

$\phi\rightarrow\psi \mapsto \neg\phi'\wedge\psi'$ *

$\exists x.\phi \mapsto \forall x.\phi'$

$\forall x.\phi \mapsto \exists x.\phi'$

** NB typo in Appendix*