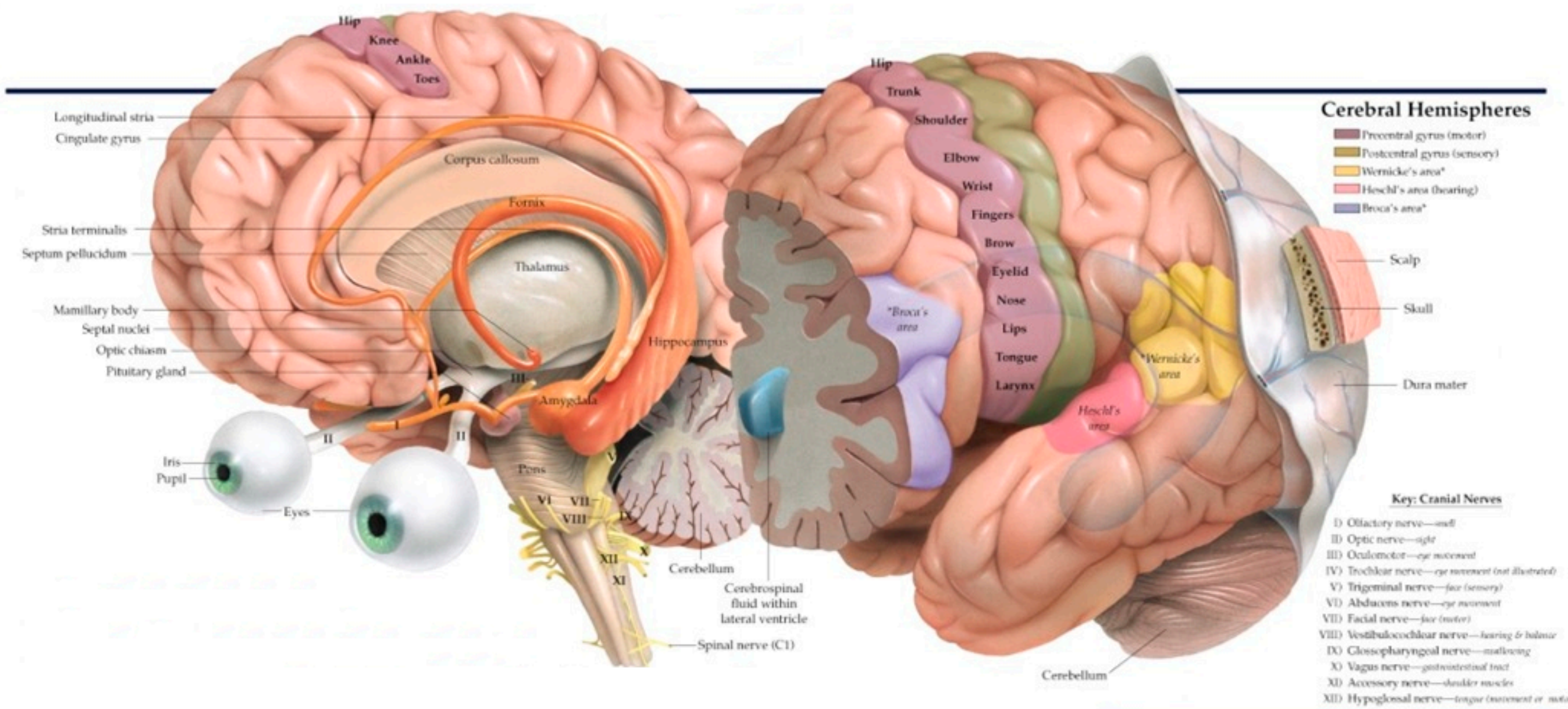
Semantic Theory
Week 9 – Distributional Formal Semantics

Noortje Venhuizen
Harm Brouwer

Universität des Saarlandes

Summer 2020

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

NATURAL LANGUAGE SEMANTICS

Model-theoretic Semantics

- Truth-conditional meaning
- Logical entailment
- Compositionality

?

Distributional Semantics

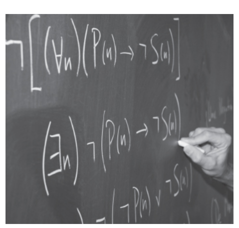
- Semantic similarity
- Empirically driven
- Cognitively inspired

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

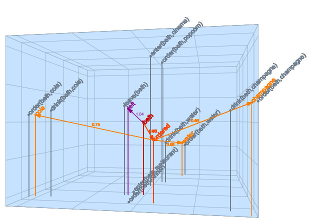
A FRAMEWORK FOR DISTRIBUTIONAL FORMAL SEMANTICS

A meaning space for distributional formal semantics

Sampling a meaning space



Formal properties of the meaning space



Incremental meaning construction

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

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)

CAPTURING THE STRUCTURE OF THE WORLD

"Will rides a bike"

Will is (likely) outside

Will is not asleep

If its dark, his light is on

The bike has wheels

etc.

- Top-down world knowledge restricts propositional co-occurrence
- Two types of world knowledge constraints reflected in $S_{\mathcal{M}\times\mathcal{P}}$:
  - Individual models reflect **hard** world knowledge constraints
  - Set of models $\mathcal{M}$ reflects **probabilistic** structure of the world

SAMPLI

Incremen ... set of
propositi ... a model
(*Light Wo* ... *rld*)

- *a propos* ... *orld*
  - *p is c*
  - *addi*
- *a propos* ... *orld*
  - *p is c*
  - *addi* ... *ark World*
- *if p cann* ... *ed to the Light W*

```
/*
 * Copyright 2017-2020 Harm Brouwer <me@hbrouwer.eu>
 *     and Noortje Venhuizen <njvenhuizen@gmail.com>
 *
 * Licensed under the Apache License, Version 2.0 (the "License");
 * you may not use this file except in compliance with the License.
 * You may obtain a copy of the License at
 *
 *     http://www.apache.org/licenses/LICENSE-2.0
 *
 * Unless required by applicable law or agreed to in writing, software
 * distributed under the License is distributed on an "AS IS" BASIS,
 * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 * See the License for the specific language governing permissions and
 * limitations under the License.
 */

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% "Link, it is I, Sahasrahla. I am communicating to you across the void   %
% through telepathy... The place where you now stand was the Golden Land, %
% but evil power turned it into the Dark World. The wizard has broken the %
% wise men's seal and opened a gate to link the worlds at Hyrule Castle.  %
% In order to save this half of the world, the Light World, you must win  %
% back the Golden Power."                                                 %
%       — Sahasrahla                                                      %
%                                                                         %
% From: The Legend of Zelda: A Link to the Past (1992)                    %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

:- module(dfs_sampling,
        [
                op(900,fx, @+),         %% constant
                op(900,fx, @*),         %% property
                op(900,fx, @#),         %% constraint
                op(900,xfx,<-),         %% probability

                dfs_sample_models/2,
                dfs_sample_model/1,

                dfs_sample_models_mt/3
        ]).

:- use_module(library(debug)). % topic: dfs_sampling
:- use_module(library(lists)).
:- use_module(library(ordsets)).
:- use_module(library(random)).

:- use_module(dfs_interpretation).
:- use_module(dfs_io).
:- use_module(dfs_logic).

:- public
        (@+)/1,
NORMAL  +0 ~0 -0 ⎇ master  dfs_sampling.pl          prolog  utf-8[unix]  0% ☰  1/558 ln : 1  ☰ [91]trailing
[0] 1:nvim* 2:zsh- 3:zsh  4:zsh                     "Harms-MacBook-Pro.loc" 09:02 30-Jun-20
```

SAMPLING A MEANING SPACE: EXAMPLE

Truth constraint: LW ⊨ *All boys*$_{\{dustin,lucas,mike\}}$ *ride a bicycle*

? Mike rides a bicycle

Dustin 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*$_{\{dustin,lucas,mike\}}$ *ride a bicycle*

✓ *Mike rides a bicycle*

Dustin rides a bicycle

Mike rides a bicycle

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

DFS MEANING SPACE

Based on a set of propositions $\mathcal{P}$, we sample a set of k models $\mathcal{M}$ which together 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	…
M_5	0	1	0	0	…
…	…	…	…	…	…

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

where $v_i(p_j) = 1$ *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 represent compositional meaning

- Standard logical operators interpreted as in model-theory

  $v_i(\neg p) = 1$ *iff* $M_i \not\models p$

  $v_i(p \wedge q) = 1$ *iff* $M_i \models p$ *and* $M_i \models q$

  etc.

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

  $v_i(\forall x\varphi) = 1$ *iff* $M_i \models \varphi[x\backslash e_1] \wedge \dots \wedge \varphi[x\backslash e_m]$

  $v_i(\exists x\varphi) = 1$ *iff* $M_i \models \varphi[x\backslash e_1] \vee \dots \vee \varphi[x\backslash e_m]$

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 φ 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 ψ given φ

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

	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 ψ given φ

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

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

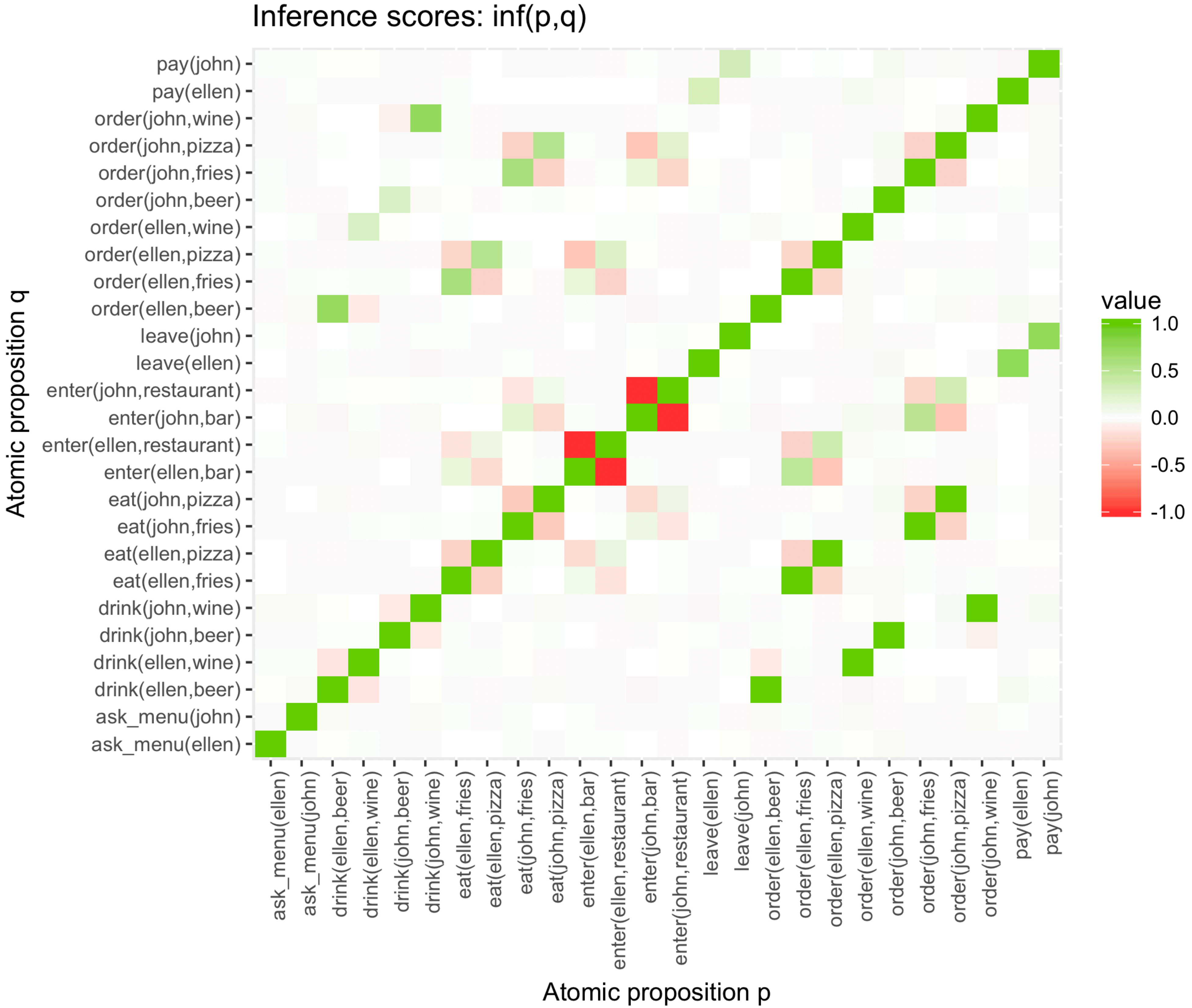
$$inf(\psi,\varphi) = 1 \Leftrightarrow \varphi \vDash \psi$$

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

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

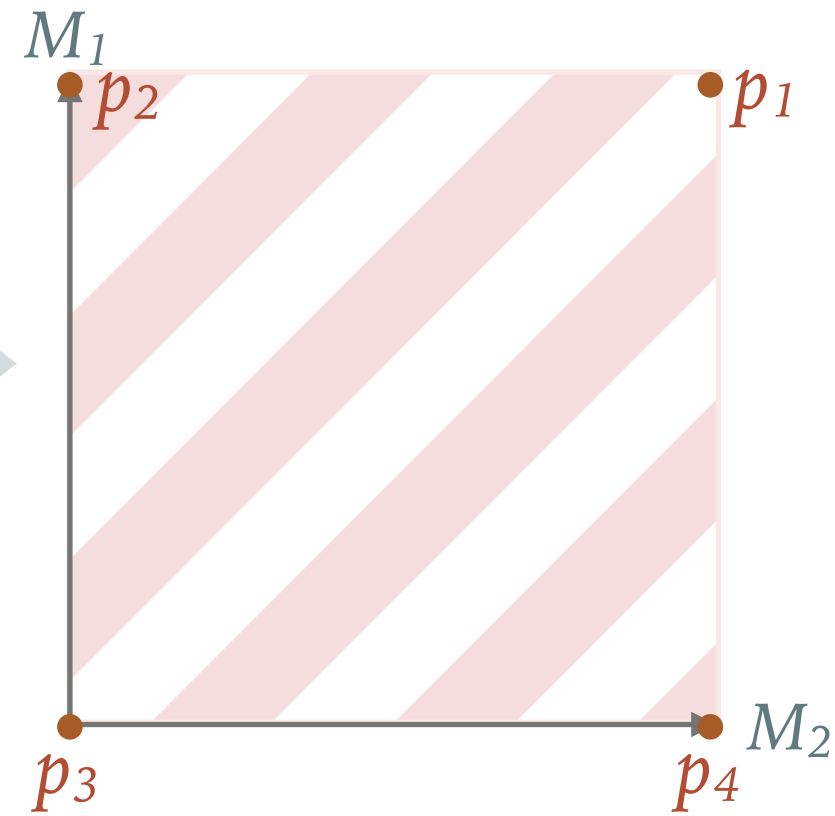
Frank et al. (2003)
Cognitive Sci.

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 lecture!**

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!