A computational model of S-selection

Aaron Steven White [1, 2] Kyle Rawlins [1]

Semantics and Linguistic Theory 26
University of Texas, Austin
14th May, 2016

Johns Hopkins University
[1]Department of Cognitive Science
[2]Center for Language and Speech Processing
[2]Science of Learning Institute

Slides available at **aswhite.net**

Introduction

Preliminary

Traditional distributional analyses have had tremendous success in helping us understand S(emantic)-selection

Preliminary

Traditional distributional analyses have had tremendous success in helping us understand S(emantic)-selection

S-selection

What type signatures does a predicate's denotation have?

Introduction

Preliminary

Traditional distributional analyses have had tremendous success in helping us understand S(emantic)-selection

S-selection

What type signatures does a predicate's denotation have?

Challenge

These analyses can be difficult to scale to an entire lexicon

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Introduction

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rules, and lexical idiosyncrasy at Marr's (1982) computational level

Introduction

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rules, and lexical idiosyncrasy at Marr's (1982) computational level
2. Collect data on ∼1000 verbs' syntactic distributions

Introduction

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rules, and lexical idiosyncrasy at Marr's (1982) computational level
2. Collect data on ∼1000 verbs' syntactic distributions
3. Given syntactic distribution data, use computational techniques to automate inference of projection rules and verbs' semantic type, controlling for lexical idiosyncrasy

Focus

Clause-embedding predicates ($\sim$1000 in English)

Introduction

Focus

Clause-embedding predicates ($\sim$1000 in English)

Case study

Responsive predicates: take both interrogative and declaratives

(1) John knows {that, whether} it's raining.

Focus

Clause-embedding predicates (∼1000 in English)

Case study

Responsive predicates: take both interrogative and declaratives

(1) John knows {that, whether} it's raining.

Importance

Deep literature on S-selection properties of responsives

Do they take questions, propositions, or both? (Karttunen 1977, Groenendijk & Stokhof 1984, Heim 1994, Ginzburg 1995, Lahiri 2002, George 2011, Rawlins 2013, Spector & Egre 2015, Uegaki 2015)

Outline

Selection and clausal embedding

Multiplicity

Many verbs are syntactically multiplicitous

(2) a. John knows {that, whether} it's raining.
    b. John wants {it to rain, rain}.

Multiplicity

Many verbs are syntactically multiplicitous

(2) a. John knows {that, whether} it's raining.
    b. John wants {it to rain, rain}.

Syntactic multiplicity does not imply semantic multiplicity

(3) a. John knows [what the answer is]$_S$.
    b. John knows [the answer]$_{NP}$.

Multiplicity

Many verbs are **syntactically multiplicitous**

(2) a. John knows {that, whether} it's raining.
    b. John wants {it to rain, rain}.

Syntactic multiplicity does not imply **semantic multiplicity**

(3) a. John knows [what the answer is]$_S$.
    b. John knows [the answer]$_{NP}$.

$[\![(3b)]\!] = [\![(3a)]\!]$ suggests it is possible for $\textbf{type}([\![\text{NP}]\!]) = \textbf{type}([\![\text{S}]\!])$

Projection

Semantic type [___Q] (Grimshaw's notation)

Syntactic type [___S] [___NP]

Projection

$$\begin{array}{llcl} \text{Semantic type} & & [\_\_\_Q] & \text{(Grimshaw's notation)} \\ \text{Projection} & \swarrow & & \searrow \\ \text{Syntactic type} & [\_\_\_S] & & [\_\_\_NP] \end{array}$$

Projection

Semantic type	$\langle\langle\langle s,t\rangle,t\rangle, t\rangle$	(Montagovian notation)
Projection		
Syntactic type	[___S]	[___NP]

Projection

What do the projection rules look like?

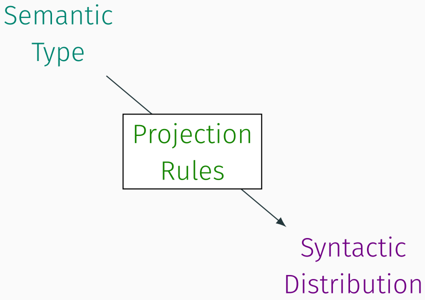
How are a verb's semantic type signatures projected onto its syntactic type signatures (subcategorization frames)? (Gruber 1965, Jackendoff 1972, Carter 1976, Grimshaw 1979, 1990, Chomsky 1981, Pesetsky 1982, 1991, Pinker 1984, 1989, Levin 1993)

Semantic type [___Q]

Projection

Syntactic type [___S] [___NP]

A model of S-selection and projection



Lexical idiosyncrasy

Observed syntactic distributions are not a perfect reflection of semantic type + projection rules

Example

Some Q(uestion)-selecting verbs allow concealed questions...

(4) a. Mary asked what time it was.
    b. Mary asked the time.

Lexical idiosyncrasy

Observed syntactic distributions are not a perfect reflection of semantic type + projection rules

Example

Some Q(uestion)-selecting verbs allow concealed questions...

(4) a. Mary asked what time it was.
b. Mary asked the time.

...others do not (Grimshaw 1979, Pesetsky 1982, 1991, Nathan 2006, Frana 2010, a.o.)

(5) a. Mary wondered what time it was.
b. *Mary wondered the time.

Two kinds of lexical idiosyncrasy

Grimshaw (1979)

Verbs are related to semantic type signatures (**S-selection**) and syntactic type signatures (**C-selection**)

Two kinds of lexical idiosyncrasy

Grimshaw (1979)

Verbs are related to semantic type signatures (**S-selection**) and syntactic type signatures (**C-selection**)

Pesetsky (1982, 1991)

Verbs are related to semantic type signatures (**S-selection**); **C-selection** is an epiphenomenon of verbs' abstract case

Two kinds of lexical idiosyncrasy

Grimshaw (1979)

Verbs are related to semantic type signatures (**S-selection**) and syntactic type signatures (**C-selection**)

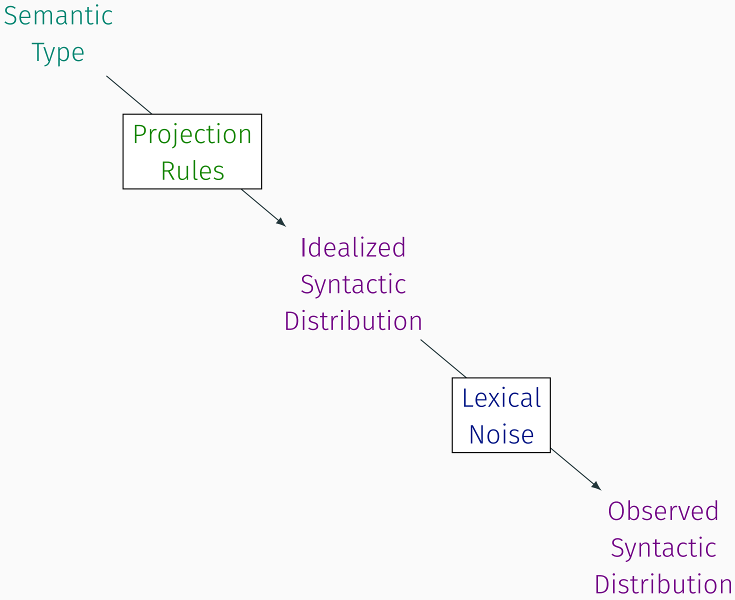
Pesetsky (1982, 1991)

Verbs are related to semantic type signatures (**S-selection**); **C-selection** is an epiphenomenon of verbs' abstract case

Shared core

Lexical noise (idiosyncrasy) alters verbs' idealized syntactic distributions

A model of S-selection and projection



Specifying the model

Question

How do we represent each object in the model?

Specifying the model

Question

How do we represent each object in the model?

A minimalistic answer

Every object is a matrix of boolean values

Specifying the model

Question

How do we represent each object in the model?

A minimalistic answer

Every object is a matrix of boolean values

Strategy

1. Give model in terms of sets and functions

Specifying the model

Question

How do we represent each object in the model?

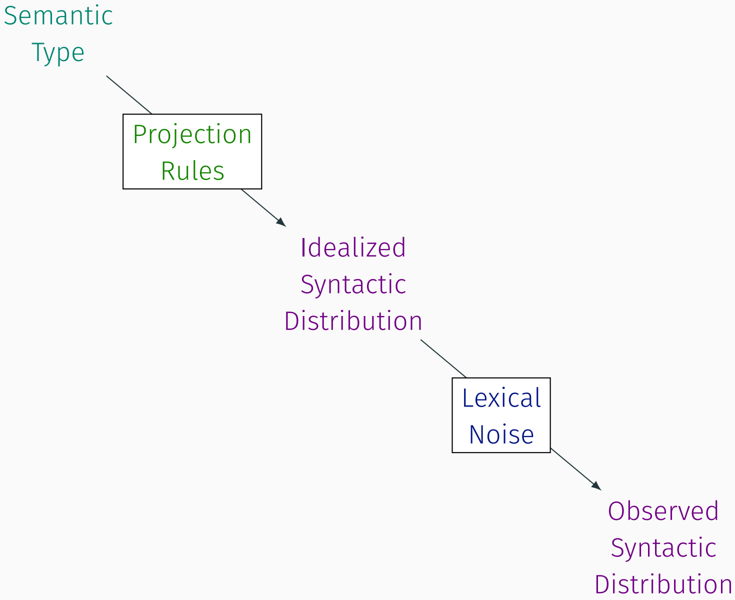
A minimalistic answer

Every object is a matrix of boolean values

Strategy

1. Give model in terms of sets and functions
2. Convert this model into a boolean matrix model

A model of S-selection and projection



A boolean model of S-selection

$$\text{know} \rightarrow \{[\_\_\_\text{P}], [\_\_\_\text{Q}]\}$$

A boolean model of S-selection

know $\rightarrow$ {[___P], [___Q]}   wonder $\rightarrow$ {[___Q]}

A boolean model of S-selection

think $\rightarrow$ {[___P]}   know $\rightarrow$ {[___P], [___Q]}   wonder $\rightarrow$ {[___Q]}

A boolean model of S-selection

think $\rightarrow$ {[___P]} know $\rightarrow$ {[___P], [___Q]} wonder $\rightarrow$ {[___Q]}

$$S = \begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{ccc} [\_\_\_P] & [\_\_\_Q] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of projection

[___P] → {[___that S], [___NP], ...}   [___Q] → {[___whether S], [___NP], ...}

A boolean model of projection

[___P] → {[___that S], [___NP], ...}   [___Q] → {[___whether S], [___NP], ...}

$$\boldsymbol{\Pi} = \begin{array}{c} \\ [\_\_\_\text{P}] \\ [\_\_\_\text{Q}] \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\_\text{that S}] & [\_\_\_\text{whether S}] & [\_\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{VERB}, \text{SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} \mathsf{S}(\text{VERB}, t) \wedge \mathbf{\Pi}(t, \text{SYNTYPE})$$

A boolean model of idealized syntactic distribution

$$\hat{\text{D}}(\text{VERB}, \text{SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} \text{S}(\text{VERB}, t) \wedge \mathbf{\Pi}(t, \text{SYNTYPE})$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{ccc} [\_\_\text{P}] & [\_\_\text{Q}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array} \qquad \begin{array}{l} \\ [\_\_\text{P}] \\ [\_\_\text{Q}] \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{VERB}, \text{SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} \mathsf{S}(\text{VERB}, t) \wedge \mathbf{\Pi}(t, \text{SYNTYPE})$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array}\begin{array}{c} \begin{array}{ccc} [\_\_\text{P}] & [\_\_\text{Q}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array} \qquad \begin{array}{l} \\ [\_\_\text{P}] \\ [\_\_\text{Q}] \\ \cdots \end{array}\begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array}\begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{know}, [\_\_\text{that S}]) = \bigvee_{t\in\{[\_\_\text{P}],[\_\_\text{Q}],\ldots\}} \mathsf{S}(\text{know}, t) \wedge \mathbf{\Pi}(t, [\_\_\text{that S}])$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array}
\begin{array}{c} \begin{array}{ccc} [__\text{P}] & [__\text{Q}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array}
\qquad
\begin{array}{l} \\ [__\text{P}] \\ [__\text{Q}] \\ \cdots \end{array}
\begin{array}{c} \begin{array}{cccc} [__\text{that S}] & [__\text{whether S}] & [__\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array}
\begin{array}{c} \begin{array}{cccc} [__\text{that S}] & [__\text{whether S}] & [__\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{know}, [\_\_\text{that S}]) = \bigvee_{t \in \{[\_\_\text{P}],[\_\_\text{Q}],\ldots\}} \mathsf{S}(\text{know}, t) \wedge \mathbf{\Pi}(t, [\_\_\text{that S}])$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{ccc} [\_\_\text{P}] & [\_\_\text{Q}] & \cdots \end{array} \\ \left(\begin{array}{ccc} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{array}\right) \end{array} \qquad \begin{array}{l} \\ [\_\_\text{P}] \\ [\_\_\text{Q}] \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \left(\begin{array}{cccc} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \left(\begin{array}{cccc} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \end{array}$$

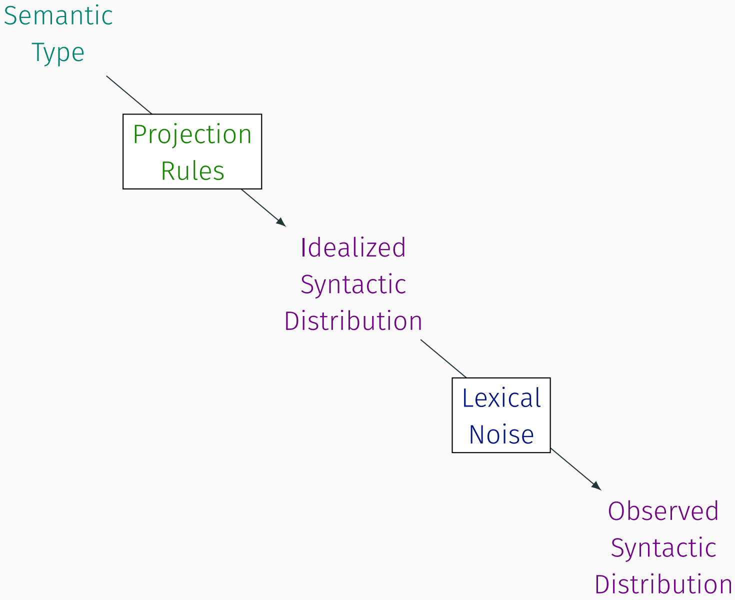
A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{wonder}, [\_\_\_\text{NP}]) = \bigvee_{t \in \{[\_\_\_\text{P}],[\_\_\_\text{Q}],\ldots\}} \mathsf{S}(\text{wonder}, t) \wedge \mathbf{\Pi}(t, [\_\_\_\text{NP}])$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array}
\begin{array}{c} \begin{array}{ccc} [___\text{P}] & [___\text{Q}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ \boxed{0 \quad 1 \quad \cdots} & & \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array}
\qquad
\begin{array}{l} \\ [___\text{P}] \\ [___\text{Q}] \\ \cdots \end{array}
\begin{array}{c} \begin{array}{cccc} [___\text{that S}] & [___\text{whether S}] & [___\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

$$\begin{array}{c|cccc} & [\_\_\_\text{that S}] & [\_\_\_\text{whether S}] & [\_\_\_\text{NP}] & \cdots \\ \hline \text{think} & 1 & 0 & 1 & \cdots \\ \text{know} & 1 & 1 & 1 & \cdots \\ \text{wonder} & 0 & 1 & \boxed{1} & \cdots \\ \cdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

A model of S-selection and projection



A boolean model of observed syntactic distribution

$$\forall t \in \mathit{SYNTYPE} : \mathtt{D}(\text{wonder}, t) = \hat{\mathtt{D}}(\text{wonder}, t) \land \mathtt{N}(\text{wonder}, t)$$

A boolean model of observed syntactic distribution

$$\forall t \in \mathit{SYNTYPE} : \mathsf{D}(\text{wonder}, t) = \hat{\mathsf{D}}(\text{wonder}, t) \wedge \mathsf{N}(\text{wonder}, t)$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array} \quad \begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of observed syntactic distribution

$$\forall t \in \mathit{SYNTYPE} : \mathsf{D}(\text{wonder}, t) = \hat{\mathsf{D}}(\text{wonder}, t) \wedge \mathsf{N}(\text{wonder}, t)$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array} \qquad \begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of observed syntactic distribution

$$\forall t \in \textit{SYNTYPE} : \mathbf{D}(\text{wonder}, t) = \hat{\mathbf{D}}(\text{wonder}, t) \land \mathbf{N}(\text{wonder}, t)$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array}
\begin{array}{c} \begin{array}{cccc} [__\text{that S}] & [__\text{whether S}] & [__\text{NP}] & \cdots \end{array} \\ \left(\begin{array}{cccc} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \end{array}
\qquad
\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array}
\begin{array}{c} \begin{array}{cccc} [__\text{that S}] & [__\text{whether S}] & [__\text{NP}] & \cdots \end{array} \\ \left(\begin{array}{cccc} 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \vdots \end{array}
\begin{array}{c} \begin{array}{cccc} [__\text{that S}] & [__\text{whether S}] & [__\text{NP}] & \cdots \end{array} \\ \left(\begin{array}{cccc} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}\right) \end{array}$$

Question

What is this model useful for?

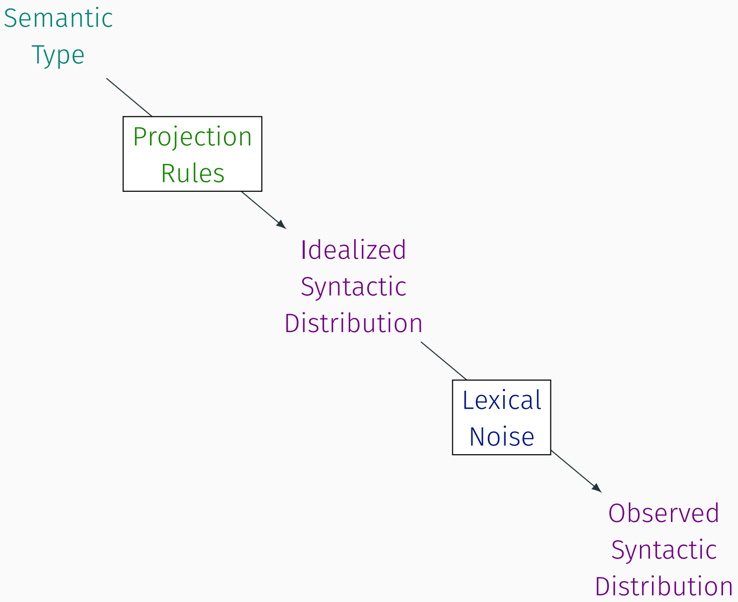
Answer

In conjunction with modern computational techniques, this model allow us to scale distributional analysis to an entire lexicon

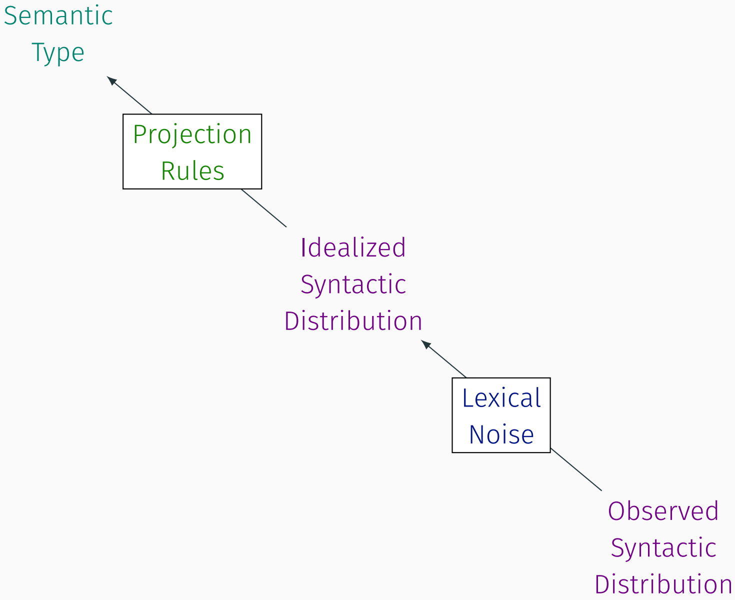
Basic idea

Distributional analysis corresponds to reversing model arrows

A model of S-selection and projection



A model of S-selection and projection



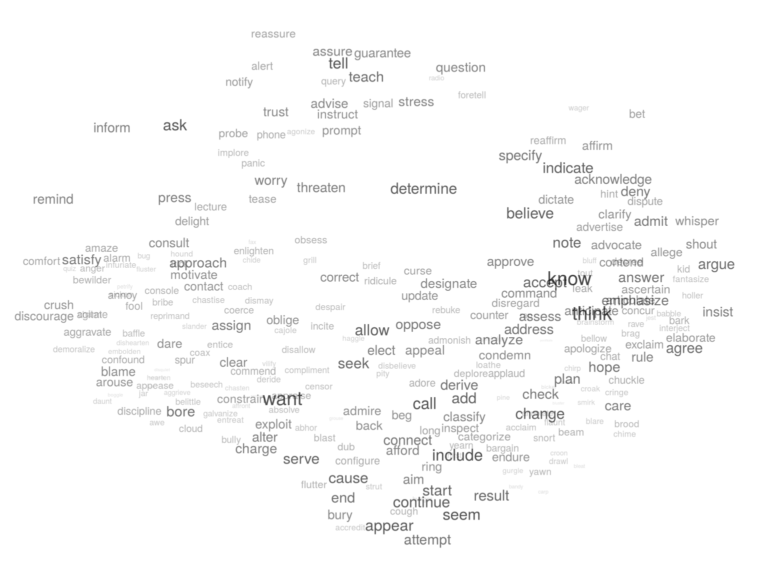
The MegaAttitude data set

Ordinal (1-7 scale) acceptability ratings

MegaAttitude materials

Ordinal (1-7 scale) acceptability ratings
for
1000 clause-embedding verbs

Verb selection



Ordinal (1-7 scale) acceptability ratings

for

1000 clause-embedding verbs

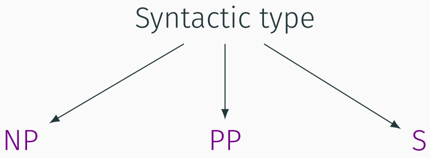
$\times$

50 syntactic frames

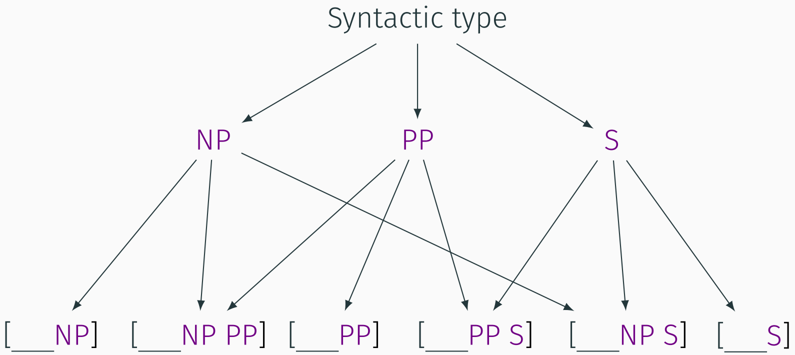
Sentence construction

Challenge

Automate construction of a very large set of frames in a way that is sufficiently general to many verbs


Syntactic type
NP
PP
S

Frame construction

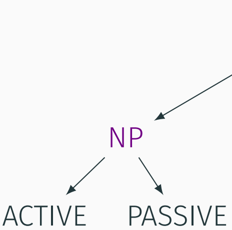


Frame construction

Syntactic type

NP PP S

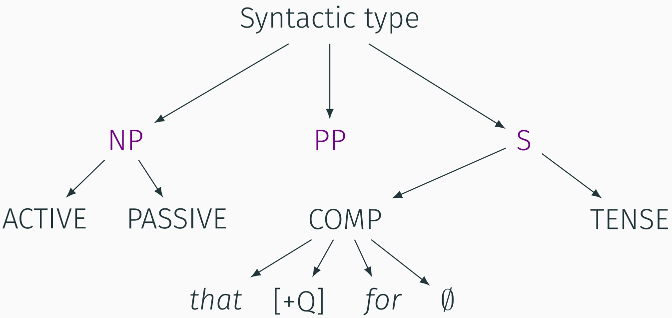
Frame construction



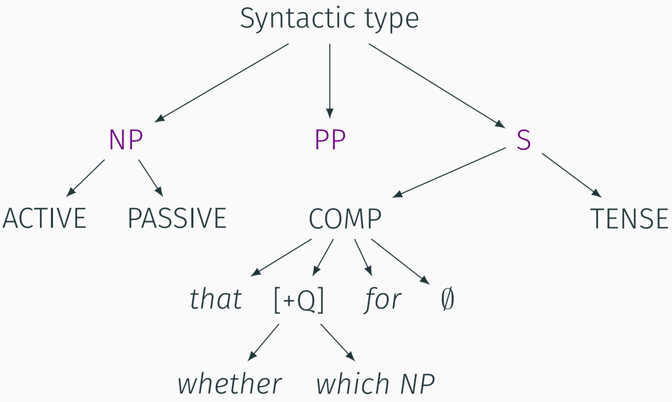
Frame construction

- Syntactic type
  - NP
    - ACTIVE
    - PASSIVE
  - PP
  - S
    - COMP
    - TENSE

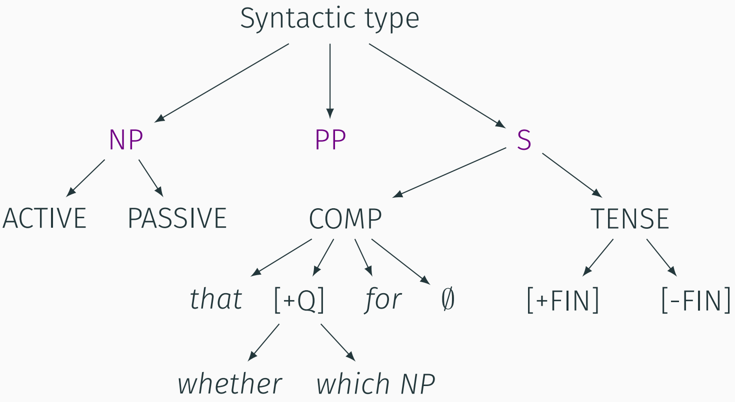
Frame construction



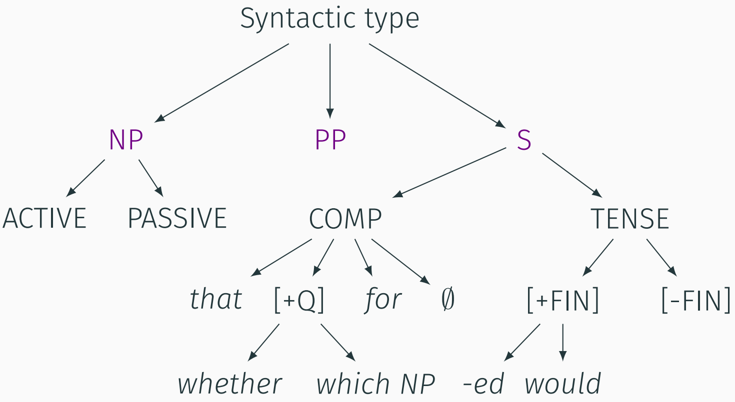
Frame construction



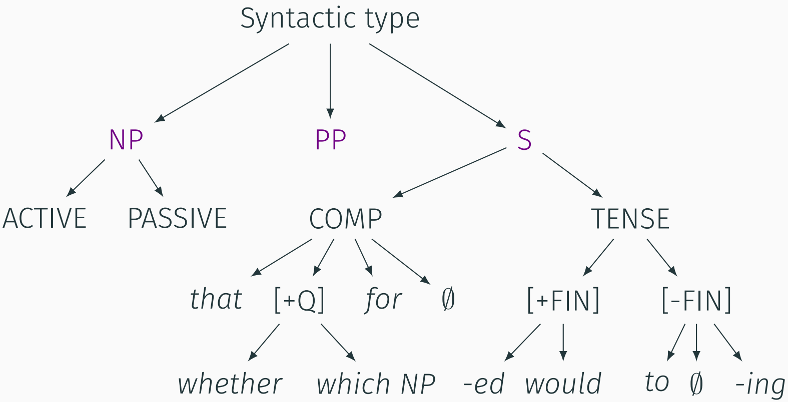
Frame construction



Frame construction



Frame construction



Sentence construction

Challenge

Automate construction of a very large set of frames in a way that is sufficiently general to many verbs

Solution

Construct semantically bleached frames using indefinites

Sentence construction

Challenge

Automate construction of a very large set of frames in a way that is sufficiently general to many verbs

Solution

Construct semantically bleached frames using indefinites

(6) Examples of responsives

a. *know* + **NP V {that, whether} S**
Someone knew {that, whether} something happened.

Challenge

Automate construction of a very large set of frames in a way that is sufficiently general to many verbs

Solution

Construct semantically bleached frames using indefinites

(6) Examples of responsives

a. *know* + NP V {that, whether} S
Someone knew {that, whether} something happened.

b. *tell* + NP V NP {that, whether} S
Someone told someone {that, whether} something happened.

Sentence construction

Challenge

Automate construction of a very large set of frames in a way that is sufficiently general to many verbs

Solution

Construct semantically bleached frames using indefinites

(6) Examples of responsives

a. *know* + **NP V {that, whether} S**
Someone knew {that, whether} something happened.

b. *tell* + **NP V NP {that, whether} S**
Someone told someone {that, whether} something happened.

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
    - Each verb only once per list

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
  - Each verb only once per list
  - Each frame only once per list

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
  - Each verb only once per list
  - Each frame only once per list
- 727 unique Mechanical Turk participants

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
  - Each verb only once per list
  - Each frame only once per list
- 727 unique Mechanical Turk participants
  - Annotators allowed to do multiple lists, but never the same list twice

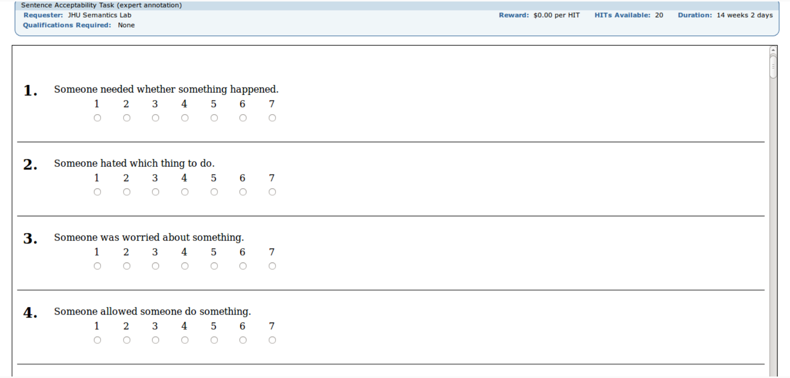
Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
    - Each verb only once per list
    - Each frame only once per list
- 727 unique Mechanical Turk participants
    - Annotators allowed to do multiple lists, but never the same list twice
- 5 judgments per item

Data collection

- 1,000 verbs $\times$ 50 syntactic frames = 50,000 sentences
- 1,000 lists of 50 items each
    - Each verb only once per list
    - Each frame only once per list
- 727 unique Mechanical Turk participants
    - Annotators allowed to do multiple lists, but never the same list twice
- 5 judgments per item
    - No annotator sees the same sentence more than once

Task



Sentence Acceptability Task (expert annotation)

Requester: JHU Semantics Lab **Reward:** $0.00 per HIT **HITs Available:** 20 **Duration:** 14 weeks 2 days

Qualifications Required: None

1. Someone needed whether something happened.

1 2 3 4 5 6 7

2. Someone hated which thing to do.

1 2 3 4 5 6 7

3. Someone was worried about something.

1 2 3 4 5 6 7

4. Someone allowed someone do something.

1 2 3 4 5 6 7

Turktools (Erlewine & Kotek 2015)

Validating the data

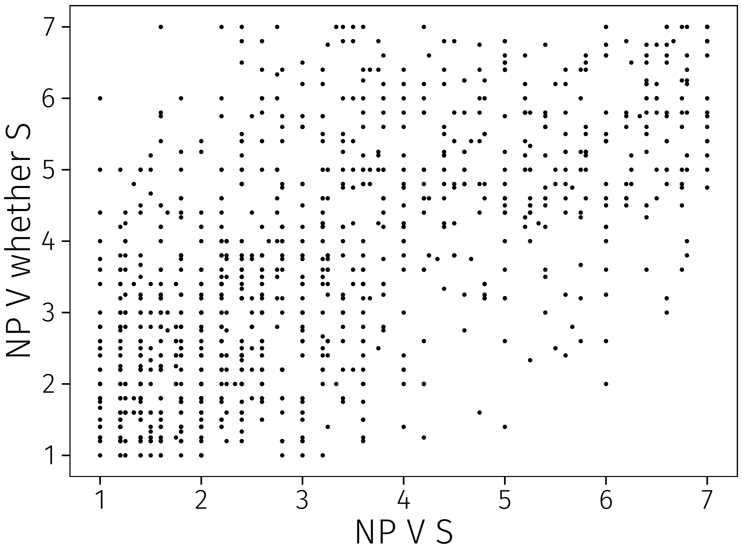
Interannotator agreement

Spearman rank correlation calculated by list on a pilot 30 verbs

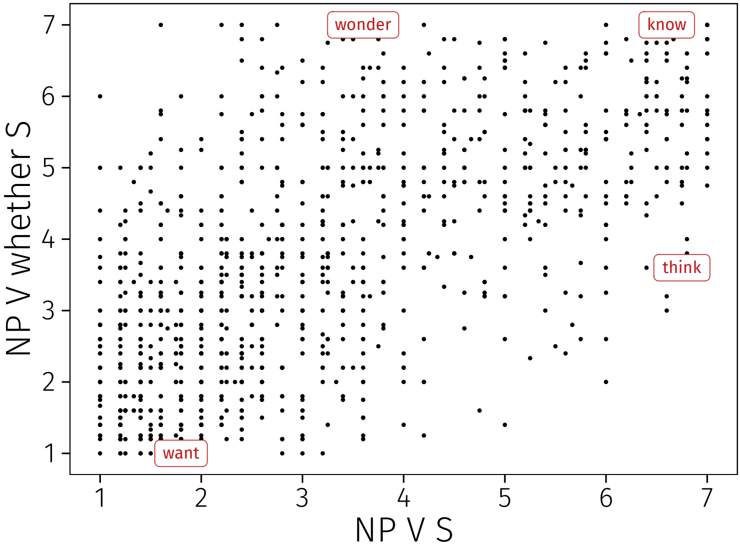
Pilot verb selection

Same verbs used by White (2015), White et al. (2015), selected based on Hacquard & Wellwood's (2012) attitude verb classification

1. **Linguist-to-linguist**
   median: 0.70, *95% CI:* [0.62, 0.78]
2. **Linguist-to-annotator**
   median: 0.55, *95% CI:* [0.52, 0.58]
3. **Annotator-to-annotator**
   median: 0.56, *95% CI:* [0.53, 0.59]

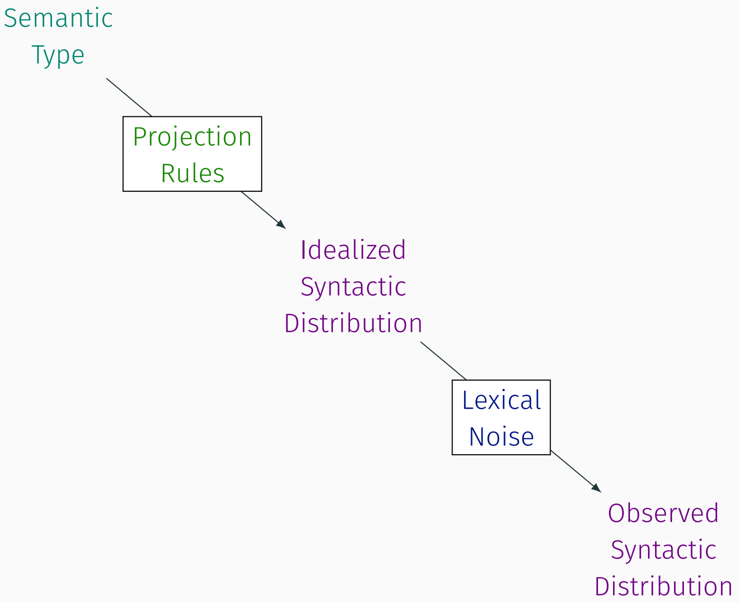

NP V whether S
NP V S
1
2
3
4
5
6
7

Results

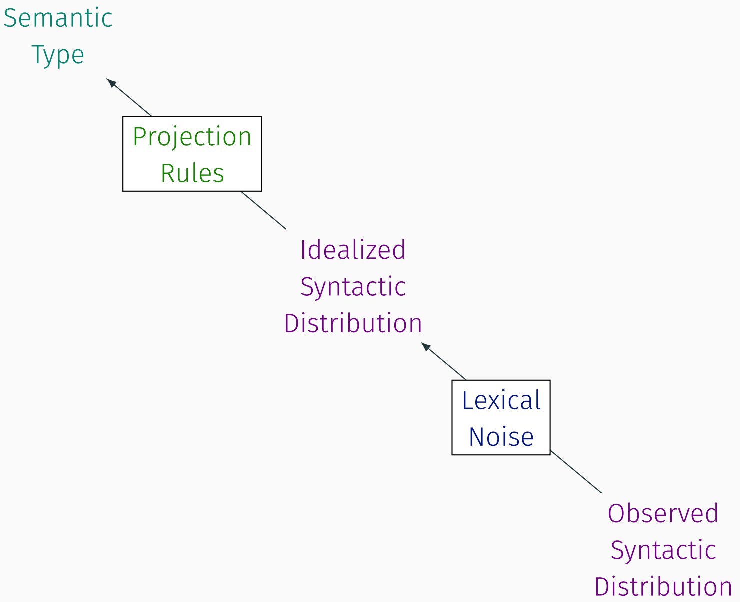


Model fitting and results

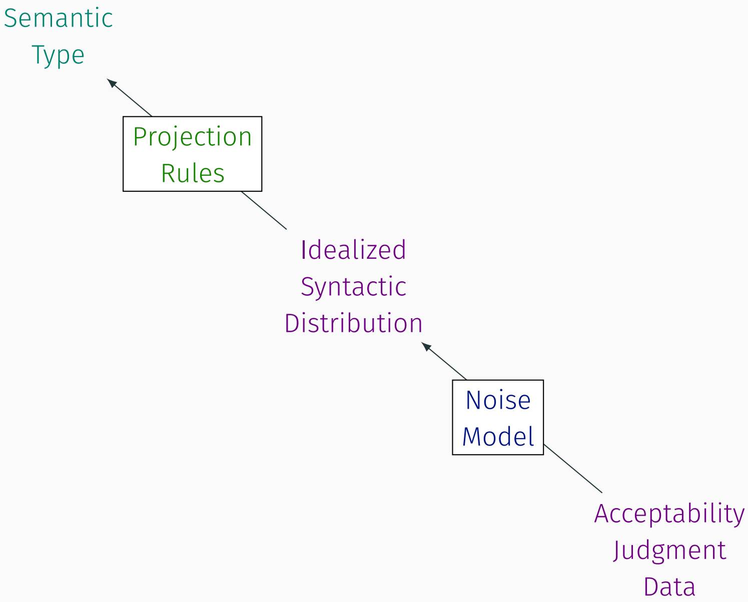
A model of S-selection and projection



A model of S-selection and projection



A model of S-selection and projection



Goal

Find representations of verbs' semantic type signatures and projection rules that best explain the acceptability judgments

Fitting the model

Goal

Find representations of verbs' semantic type signatures and projection rules that best explain the acceptability judgments

Challenges

1. Infeasible to search over $2^{1000T} \times 2^{50T}$ possible configurations ($T = \#$ of type signatures)
2. Finding the best boolean model fails to capture uncertainty inherent in judgment data

Solution

Search probability distributions over verbs' semantic type signatures and projection rules

Solution

Search probability distributions over verbs' semantic type signatures and projection rules

Going probabilistic

Wrap boolean expressions in probability measures

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\mathsf{VERB}, \mathsf{SYNTYPE}) = \bigvee_{t \in \mathsf{SEMTYPES}} \mathsf{S}(\mathsf{VERB}, t) \wedge \mathbf{\Pi}(t, \mathsf{SYNTYPE})$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{ccc} [\_\_\text{P}] & [\_\_\text{Q}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array} \qquad \begin{array}{l} \\ [\_\_\text{P}] \\ [\_\_\text{Q}] \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

$$\begin{array}{l} \\ \text{think} \\ \text{know} \\ \text{wonder} \\ \cdots \end{array} \begin{array}{c} \begin{array}{cccc} [\_\_\text{that S}] & [\_\_\text{whether S}] & [\_\_\text{NP}] & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of idealized syntactic distribution

$$\hat{\mathsf{D}}(\text{know}, [\_\_\text{that S}]) = 1 - \prod_{t \in \{[\_\_\text{P}],[\_\_\text{Q}],\ldots\}} 1 - \mathsf{S}(\text{know}, t) \times \boldsymbol{\Pi}(t, [\_\_\text{that S}])$$

$$\begin{array}{c c}
 & \begin{array}{ccc} [__\text{P}] & [__\text{Q}] & \cdots \end{array} \\
\begin{array}{l} \text{think} \\ \text{know} \\ \text{wonder} \\ \ldots \end{array} & \begin{pmatrix} 0.94 & 0.03 & \cdots \\ 0.97 & 0.91 & \cdots \\ 0.17 & 0.93 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}
\end{array}
\qquad
\begin{array}{c c}
 & \begin{array}{ccc} [__\text{that S}] & [__\text{whether S}] & \cdots \end{array} \\
\begin{array}{l} [__\text{P}] \\ [__\text{Q}] \\ \ldots \end{array} & \begin{pmatrix} 0.99 & 0.12 & \cdots \\ 0.07 & 0.98 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}
\end{array}$$

$$\begin{array}{c c}
 & \begin{array}{ccc} [__\text{that S}] & [__\text{whether S}] & \cdots \end{array} \\
\begin{array}{l} \text{think} \\ \text{know} \\ \text{wonder} \\ \ldots \end{array} & \begin{pmatrix} 0.97 & 0.14 & \cdots \\ 0.95 & 0.99 & \cdots \\ 0.12 & 0.99 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}
\end{array}$$

Wrapping with probabilities

$$
\begin{aligned}
\mathbb{P}(\mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}]) &= \mathbb{P}(\mathsf{S}[\mathsf{VERB}, t])\mathbb{P}(\mathbf{\Pi}[t, \mathsf{SYNTYPE}] \mid \mathsf{S}[\mathsf{VERB}, t]) \\
&= \mathbb{P}(\mathsf{S}[\mathsf{VERB}, t])\mathbb{P}(\mathbf{\Pi}[t, \mathsf{SYNTYPE}]) \\
\mathbb{P}\left(\bigvee_t \mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}]\right) &= \mathbb{P}\left(\neg \bigwedge_t \neg(\mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}])\right) \\
&= 1 - \mathbb{P}\left(\bigwedge_t \neg(\mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}])\right) \\
&= 1 - \prod_t \mathbb{P}\left(\neg(\mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}])\right) \\
&= 1 - \prod_t 1 - \mathbb{P}\left(\mathsf{S}[\mathsf{VERB}, t] \wedge \mathbf{\Pi}[t, \mathsf{SYNTYPE}]\right) \\
&= 1 - \prod_t 1 - \mathbb{P}(\mathsf{S}[\mathsf{VERB}, t])\mathbb{P}(\mathbf{\Pi}[t, \mathsf{SYNTYPE}])
\end{aligned}
$$

Fitting the model

Algorithm

Projected gradient descent with adaptive gradient (Duchi et al. 2011)

Algorithm

Projected gradient descent with adaptive gradient (Duchi et al. 2011)

Remaining challenge

Don't know the number of type signatures T

Fitting the model

Algorithm

Projected gradient descent with adaptive gradient (Duchi et al. 2011)

Remaining challenge

Don't know the number of type signatures T

Standard solution

Fit the model with many type signatures and compare using an information criterion, e.g., the Akaike Information Criterion (AIC)

Akaike Information Criterion

High-level idea

Measures the information theoretic "distance" to the true model from the best model with T types signatures (Akaike 1974)

Akaike Information Criterion

High-level idea

Measures the information theoretic “distance” to the true model from the best model with T types signatures (Akaike 1974)

Low-level idea (cf. Gelman et al. 2013)

For each datapoint…

1. …remove that datapoint from the dataset
2. …fit the model to the remaining data
3. …predict the held-out datapoint

In the limit, model with lowest error on step 3 has lowest AIC

Fitting the model

Result

12 is the optimal number of type signatures according to AIC

Reporting findings

Remainder of talk: best model with 12 type signatures

Three findings

1. Cognitive predicates
   1.1 Two distinct type signatures [___P] and [___Q]

Findings

$[\_\_\_P]$ $[\_\_\_Q]$

Findings

$$\begin{tikzcd}[column sep=small]
\text{[___P]} \arrow[d] & \text{[___Q]} \arrow[d] \\
\text{[___that S]} & \text{[___whether S]}
\end{tikzcd}$$

Findings

Three findings

1. Cognitive predicates
    1.1 Two distinct type signatures [___P] and [___Q]
    1.2 Coercion of [___P] to [___Q] and [___Q] to [___P]

Findings

[___P] [___Q]

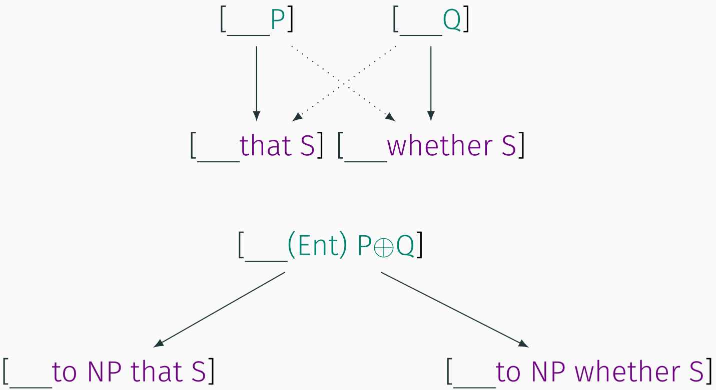
[___that S] [___whether S]

Findings

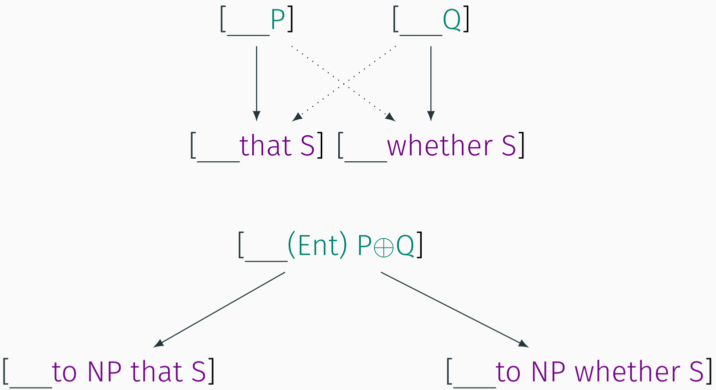
Three findings

1. Cognitive predicates
    1.1 Two distinct type signatures [___P] and [___Q]
    1.2 Coercion of [___P] to [___Q] and [___Q] to [___P]
2. Communicative predicates
    2.1 Two unified type signatures [___(Ent) P$\oplus$Q] (optional recipient) and [___Ent P$\oplus$Q] (obligatory recipient)

Findings



Findings



Hybrid types

Question

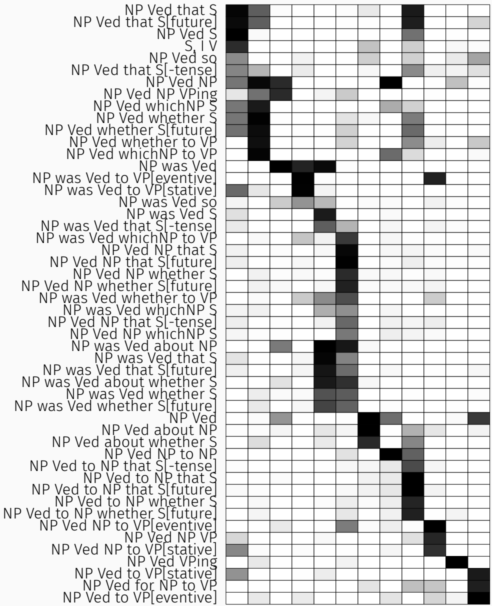
What do we mean by $P \oplus Q$?

Example

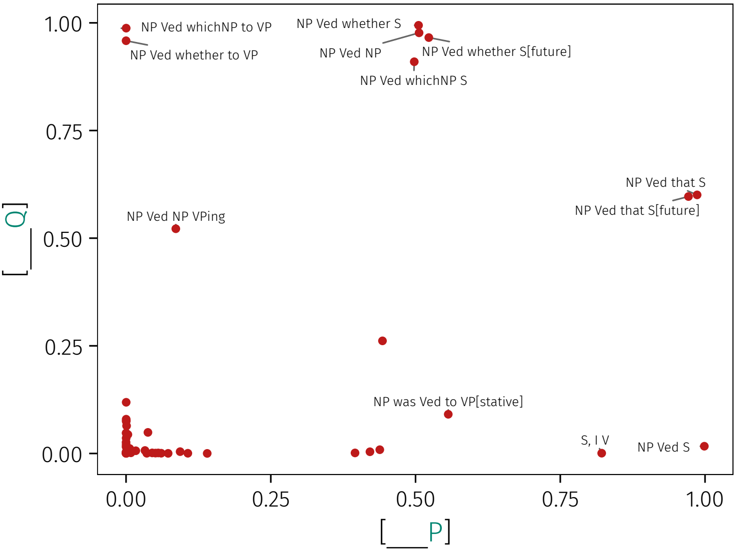
Structures with potentially both informative and inquisitive content (Groenendijk & Roelofsen 2009, a.o.)

- S-selectional behavior of responsive predicates on some accounts (Uegaki 2012; Rawlins 2013)
- Some attitudes whose content is a hybrid Lewisian (1988) subject matter (Rawlins 2013 on *think* v. *think about*)

Projection

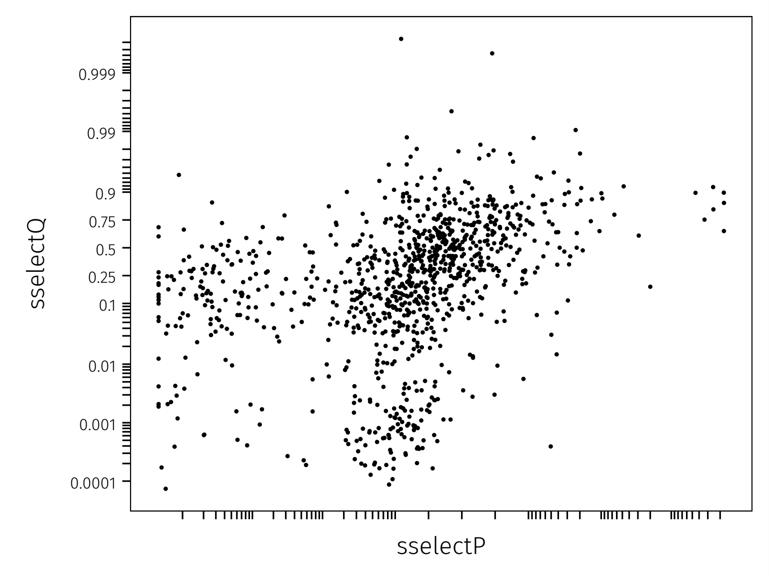

NP Ved that S
NP Ved that S[future]
NP Ved S
S, I V
NP Ved so
NP Ved that S[-tense]
NP Ved NP
NP Ved NP VPing
NP Ved whichNP S
NP Ved whether S
NP Ved whether S[future]
NP Ved whether to VP
NP Ved whichNP to VP
NP was Ved
NP was Ved to VP[eventive]
NP was Ved to VP[stative]
NP was Ved so
NP was Ved S
NP was Ved that S[-tense]
NP was Ved whichNP to VP
NP Ved NP that S
NP Ved NP that S[future]
NP Ved NP whether S
NP Ved NP whether S[future]
NP was Ved whether to VP
NP was Ved whichNP S
NP Ved NP that S[-tense]
NP Ved NP whichNP S
NP was Ved about NP
NP was Ved that S
NP was Ved that S[future]
NP was Ved about whether S
NP was Ved whether S
NP was Ved whether S[future]
NP Ved
NP Ved about NP
NP Ved about whether S
NP Ved NP to NP
NP Ved to NP that S[-tense]
NP Ved to NP that S
NP Ved to NP that S[future]
NP Ved to NP whether S
NP Ved to NP whether S[future]
NP Ved NP to VP[eventive]
NP Ved NP VP
NP Ved NP to VP[stative]
NP Ved VPing
NP Ved to VP[stative]
NP Ved for NP to VP
NP Ved to VP[eventive]

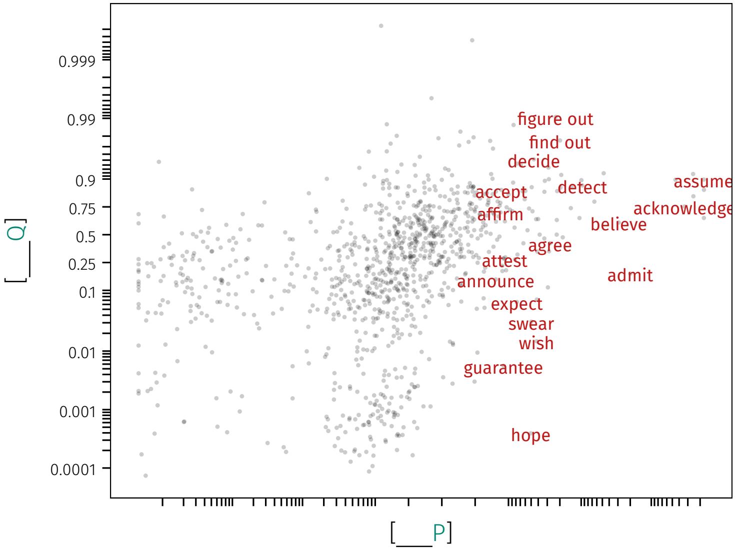
Projection

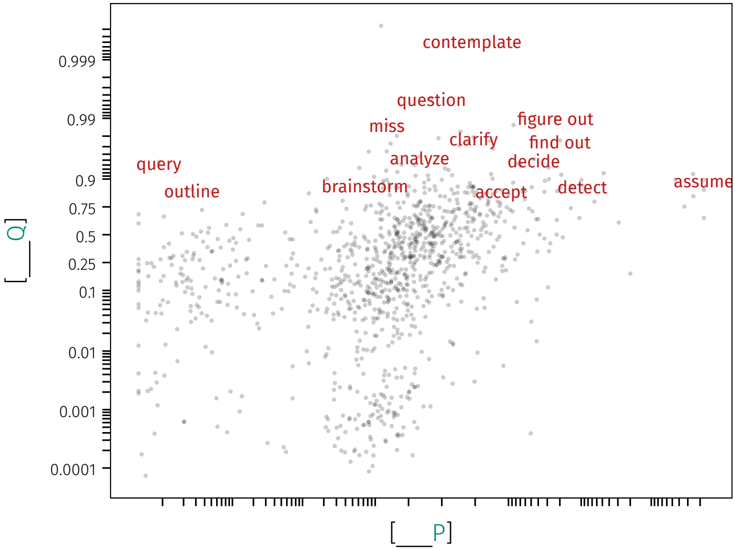


NP Ved whichNP to VP
NP Ved whether to VP
NP Ved whether S
NP Ved NP
NP Ved whether S[future]
NP Ved whichNP S
NP Ved that S
NP Ved that S[future]
NP Ved NP VPing
NP was Ved to VP[stative]
S, I V
NP Ved S
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
[___P]
[___Q]

Projection




0.999
0.99
0.9
0.75
0.5
0.25
0.1
0.01
0.001
0.0001
[___Q]
[___P]
figure out
find out
decide
detect
assume
accept
affirm
acknowledge
believe
agree
attest
announce
admit
expect
swear
wish
guarantee
hope

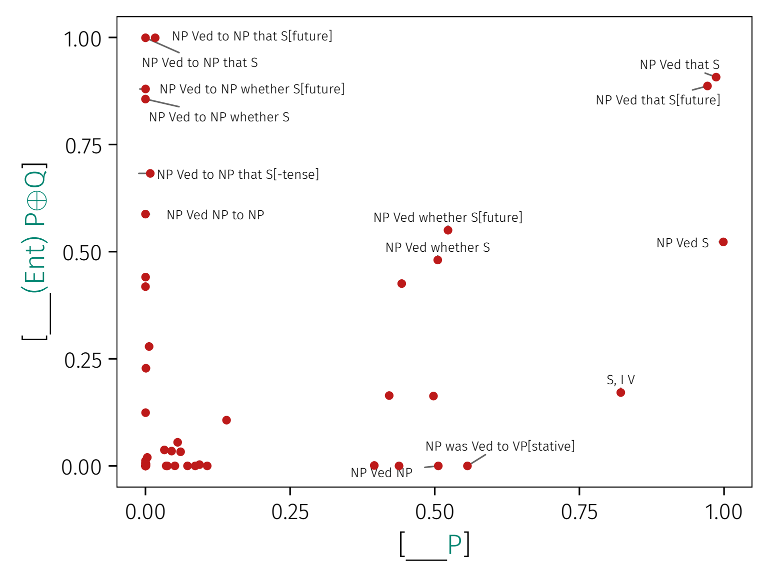


contemplate
question
figure out
miss
clarify
find out
analyze
decide
query
brainstorm
accept
detect
assume
outline
0.999
0.99
0.9
0.75
0.5
0.25
0.1
0.01
0.001
0.0001
[___Q]
[___P]

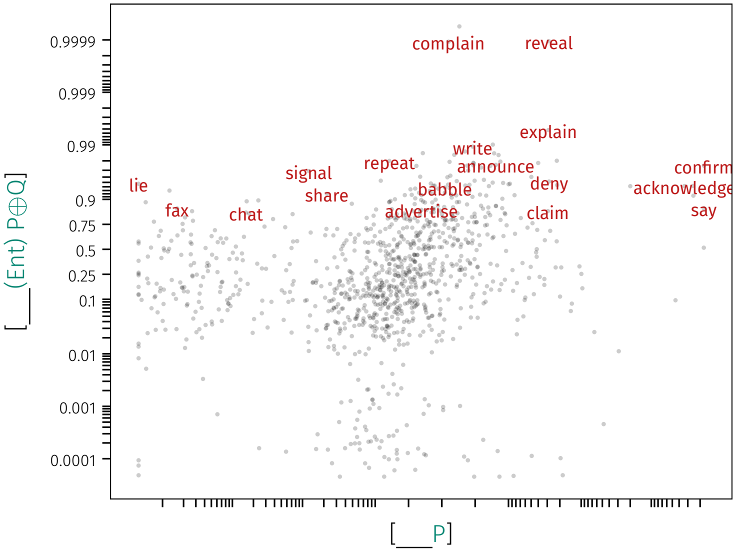
Findings

$$
\begin{tikzcd}
\lbrack \underline{}\, P \rbrack \arrow[d] \arrow[dr, dotted] & \lbrack \underline{}\, Q \rbrack \arrow[d] \arrow[dl, dotted] \\
\lbrack \underline{}\, \text{that S} \rbrack & \lbrack \underline{}\, \text{whether S} \rbrack
\end{tikzcd}
$$

Projection


NP Ved to NP that S[future]
NP Ved to NP that S
NP Ved that S
NP Ved that S[future]
NP Ved to NP whether S[future]
NP Ved to NP whether S
NP Ved to NP that S[-tense]
NP Ved NP to NP
NP Ved whether S[future]
NP Ved whether S
NP Ved S
S, I V
NP was Ved to VP[stative]
NP Ved NP
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
[___(Ent) P⊕Q]
[___P]

S-selection



0.9999
0.999
0.99
0.9
0.75
0.5
0.25
0.1
0.01
0.001
0.0001
[___(Ent) P⊕Q]
[___P]
complain
reveal
explain
write
announce
repeat
signal
share
babble
deny
advertise
claim
lie
fax
chat
confirm
acknowledge
say

Findings

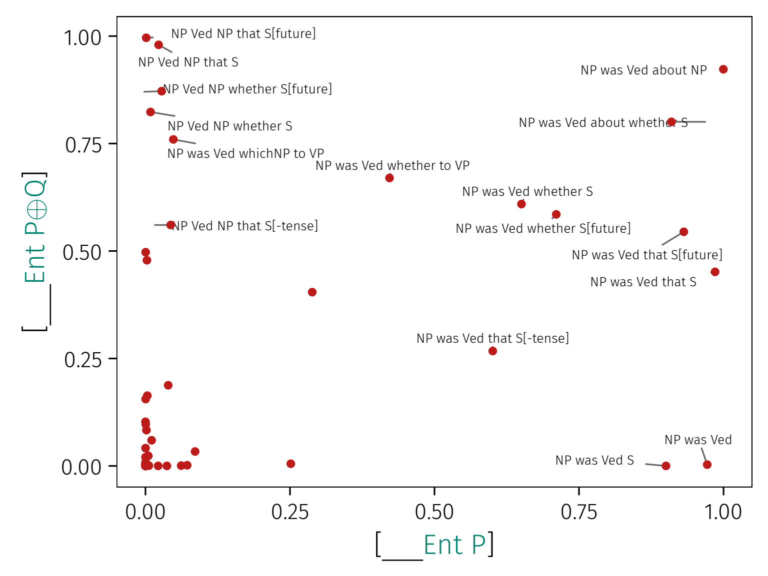
[___P] [___Q]

[___that S] [___whether S]

[___(Ent) P⊕Q]

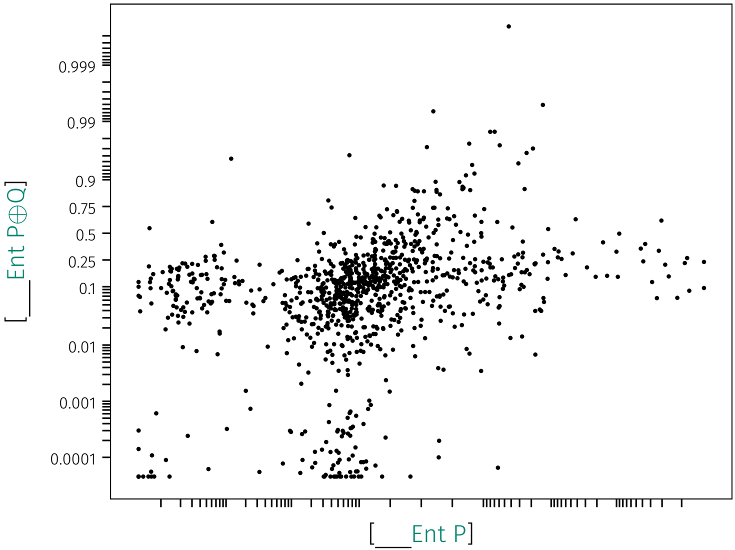
[___to NP that S] [___to NP whether S]

Projection

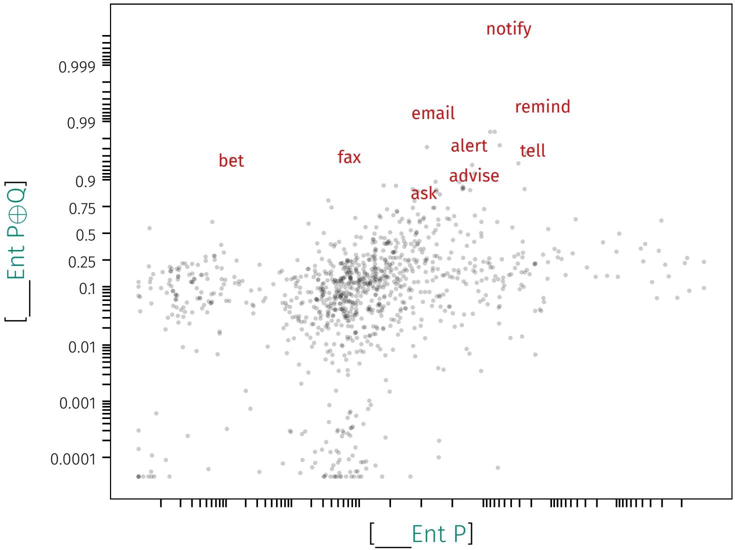


NP Ved NP that S[future]
NP Ved NP that S
NP was Ved about NP
NP Ved NP whether S[future]
NP Ved NP whether S
NP was Ved about whether S
NP was Ved whichNP to VP
NP was Ved whether to VP
NP was Ved whether S
NP Ved NP that S[-tense]
NP was Ved whether S[future]
NP was Ved that S[future]
NP was Ved that S
NP was Ved that S[-tense]
NP was Ved
NP was Ved S
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
[___Ent P⊕Q]
[___Ent P]

S-selection

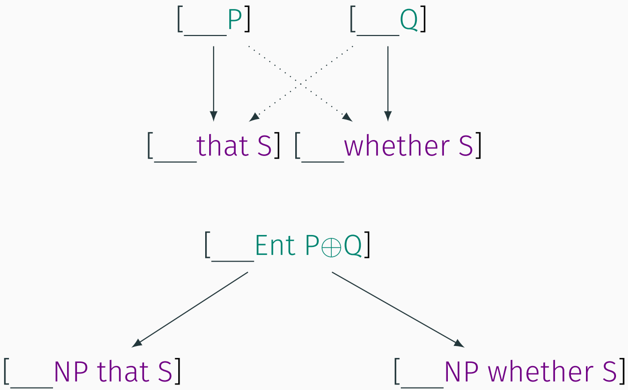


S-selection



notify
remind
email
alert
tell
bet
fax
advise
ask
0.999
0.99
0.9
0.75
0.5
0.25
0.1
0.01
0.001
0.0001
[___Ent P⊕Q]
[___Ent P]

Findings



What we conclude

Proposition and question types live alongside hybrid types, and the presence of a hybrid type correlates with communicativity

Discussion

What we conclude

Proposition and question types live alongside hybrid types, and the presence of a hybrid type correlates with communicativity

What we can exclude

Accounts that reduce (or unify) declarative and interrogative selection solely to S-selection of a single type + coercion

What we conclude

Proposition and question types live alongside hybrid types, and the presence of a hybrid type correlates with communicativity

What we can exclude

Accounts that reduce (or unify) declarative and interrogative selection solely to S-selection of a single type + coercion

Methodological point

Coercion can have measurable effects

Conclusions and future directions

Conclusions

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis

Conclusions

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Conclusions

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rule, and lexical idiosyncrasy at Marr's (1982) computational level

Conclusions

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rule, and lexical idiosyncrasy at Marr's (1982) computational level
2. Collect data on ∼1000 verbs' syntactic distributions

Conclusions

Goals

1. Demonstrate a combined experimental-computational method for scaling distributional analysis
2. Show that this method provides insight into general principles governing lexical semantic structure

Basic idea

1. Formalize S(emantic)-selection, projection rule, and lexical idiosyncrasy at Marr's (1982) computational level
2. Collect data on ∼1000 verbs' syntactic distributions
3. Given syntactic distribution data, use computational techniques to automate inference of projection rules and verbs' semantic type, controlling for lexical idiosyncrasy

Conclusions

Focus

Clause-embedding predicates ($\sim$1000 in English)

Conclusions

Focus

Clause-embedding predicates ($\sim$1000 in English)

Case study

Responsive predicates and the features that underly their selectional behavior.

(7) John knows {that, whether} it's raining.

By looking at such a large data set, we can *discover* the relevant s-selectional features, and get an angle on the problem *at the scale of the entire lexicon*.

Further investigation of type signatures

Seven other type signatures that are also remarkably coherent

Example

Many nonfinite-taking verbs

Future directions

Atomic v. structured type signatures

Currently treating type signatures as atomic but type signatures have rich structure

Example

Preliminary experiments with models that represent type structure suggest that our glosses for the types are correct

Homophony v. regular polysemy v. underspecification

Patterns in how semantic type signatures distribute across verbs may belie regular polysemy rules

Example

Preliminary experiments with a more elaborated model suggest responsive predicates display a regular polysemy (cf. George 2011)

Thanks

We are grateful to audiences at Johns Hopkins University for discussion of this work. We would like to thank Shevaun Lewis and Drew Reisinger in particular for useful comments on this talk.

This work was funded by NSF DDRIG-1456013 (*Doctoral Dissertation Research: Learning attitude verb meanings*), NSF INSPIRE BCS-1344269 (*Gradient symbolic computation*), and the JHU Science of Learning Institute.

Bibliography I

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6). 716–723.

Carter, Richard. 1976. Some linking regularities. *On Linking: Papers by Richard Carter Cambridge MA: Center for Cognitive Science, MIT (Lexicon Project Working Papers No. 25)* .

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Walter de Gruyter.

Duchi, John, Elad Hazan & Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research* 12. 2121–2159.

Bibliography II

Erlewine, Michael Yoshitaka & Hadas Kotek. 2015. A streamlined approach to online linguistic surveys. *Natural Language & Linguistic Theory* 1–15. doi:10.1007/s11049-015-9305-9. `http://link.springer.com/article/10.1007/s11049-015-9305-9`.

Frana, Ilaria. 2010. *Concealed Questions: in search of answers*: University of Massachusetts at Amherst Ph.D. dissertation.

Gelman, Andrew, Jessica Hwang & Aki Vehtari. 2013. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 1–20.

George, Benjamin Ross. 2011. *Question embedding and the semantics of answers*: University of California Los Angeles dissertation.

Ginzburg, Jonathan. 1995. Resolving questions, II. *Linguistics and Philosophy* 18(6). 567–609.

Bibliography III

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10(2). 279–326.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press.

Groenendijk, Jeroen & Floris Roelofsen. 2009. Inquisitive semantics and pragmatics. Paper presented at Stanford workshop on Language, Communication, and Rational Agency.

Groenendijk, Jeroen & Martin Stokhof. 1984. On the semantics of questions and the pragmatics of answers. *Varieties of formal semantics* 3. 143–170.

Gruber, Jeffrey Steven. 1965. *Studies in lexical relations*: Massachusetts Institute of Technology dissertation.

Hacquard, Valentine & Alexis Wellwood. 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and Pragmatics* 5(4). 1–29.

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for know. In *Proceedings of IATL*, vol. 1, 128–144.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and philosophy* 1(1). 3–44.

Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford University Press.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Lewis, David. 1988. Relevant implication. *Theoria* 54(3). 161–174.

Bibliography V

Marr, David. 1982. Vision: a computational investigation into the human representation and processing of visual information. *Henry Holt and Co.* .

Nathan, Lance Edward. 2006. *On the interpretation of concealed questions*: Massachusetts Institute of Technology dissertation.

Pesetsky, David. 1982. *Paths and categories*: MIT dissertation.

Pesetsky, David. 1991. Zero syntax: vol. 2: Infinitives.

Pinker, Steven. 1984. *Language learnability and language development*. Harvard University Press.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Rawlins, Kyle. 2013. About 'about'. In *Semantics and Linguistic Theory*, vol. 23, 336–357.

Bibliography VI

Spector, Benjamin & Paul Egre. 2015. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese* 192(6). 1729–1784.

Uegaki, Wataru. 2012. Content nouns and the semantics of question-embedding predicates. In Ana Aguilar-Guevara, Anna Chernilovskaya & Rick Nouwen (eds.), *Proceedings of SuB 16*, .

Uegaki, Wataru. 2015. *Interpreting questions under attitudes*: MIT dissertation.

White, Aaron Steven. 2015. *Information and incrementality in syntactic bootstrapping*: University of Maryland dissertation.

White, Aaron Steven, Valentine Hacquard & Jeffrey Lidz. 2015. Projecting attitudes.

Appendix

Two functions

1. Normalize for participants' judgments so they are comparable

The response model

Two functions

1. Normalize for participants' judgments so they are comparable
2. Control for **lexicosyntactic noise**

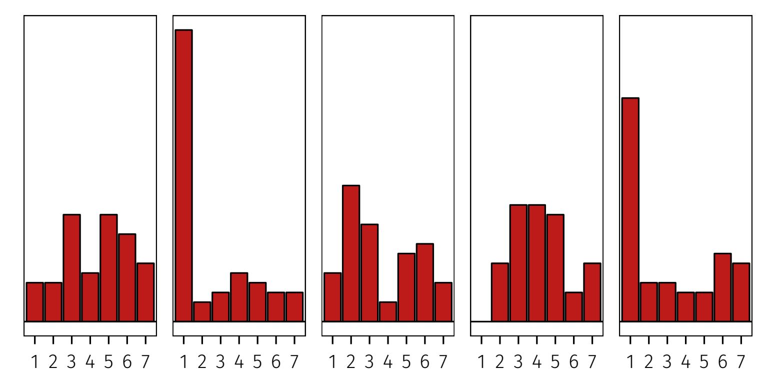
Why normalize judgments?

Necessary to control for differences in participants' use of scale

The response model

Why normalize judgments?

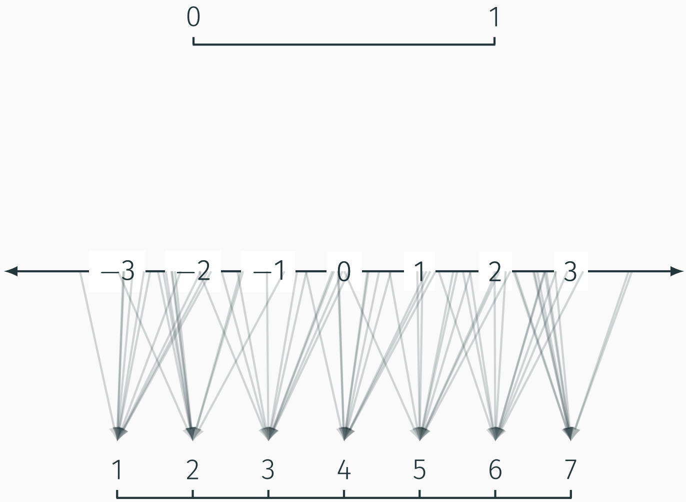
Necessary to control for differences in participants' use of scale



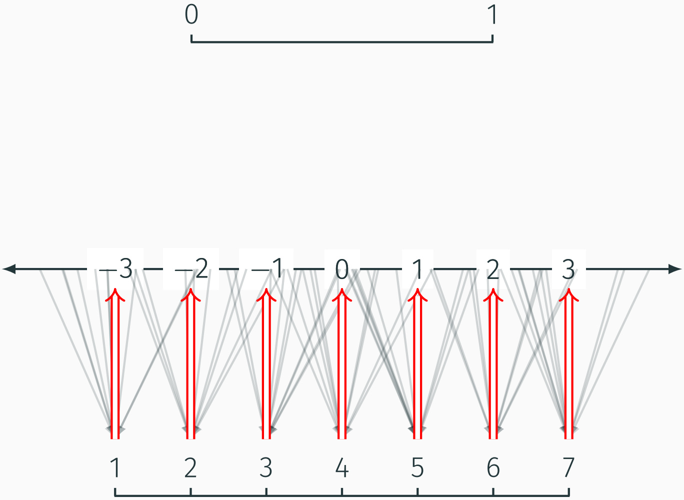
The response model

The response model

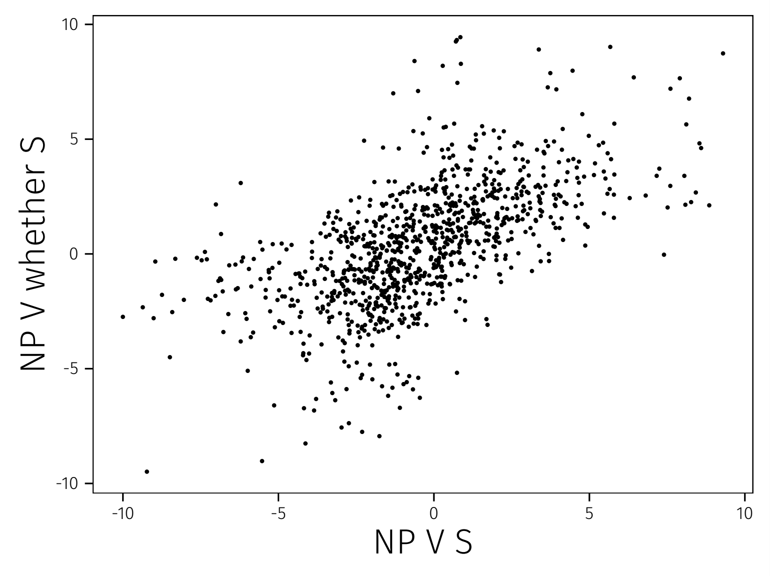
The response model



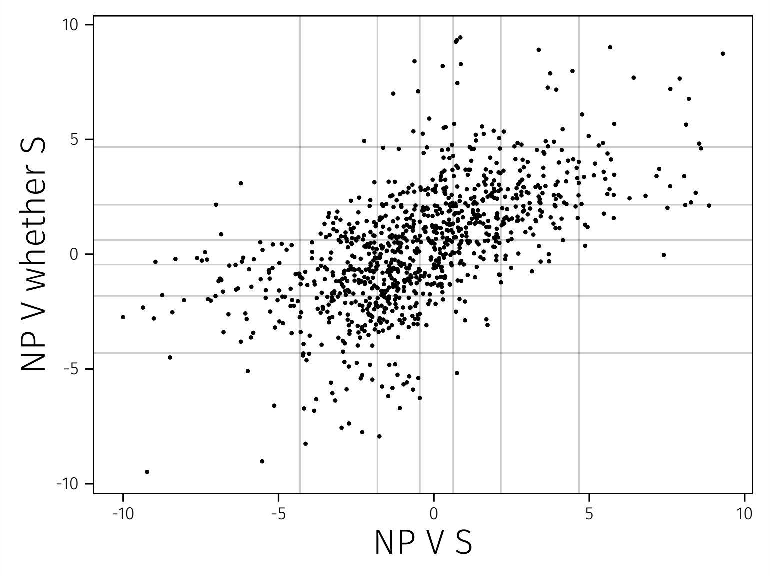
The response model

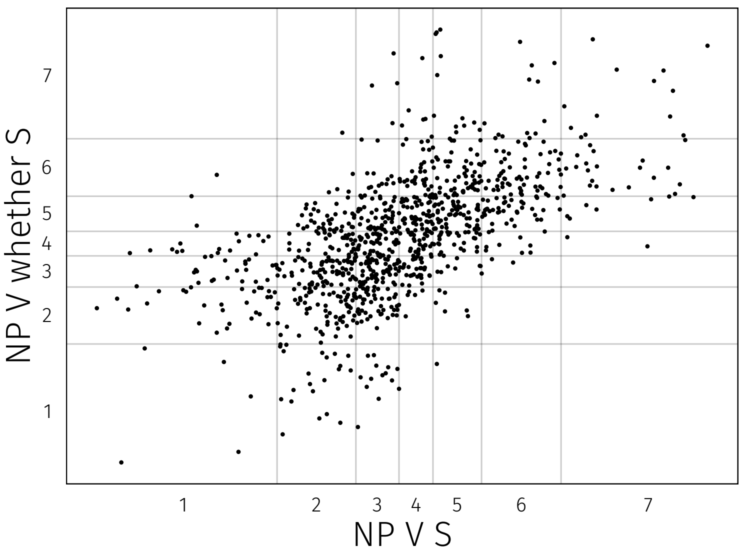


The response model

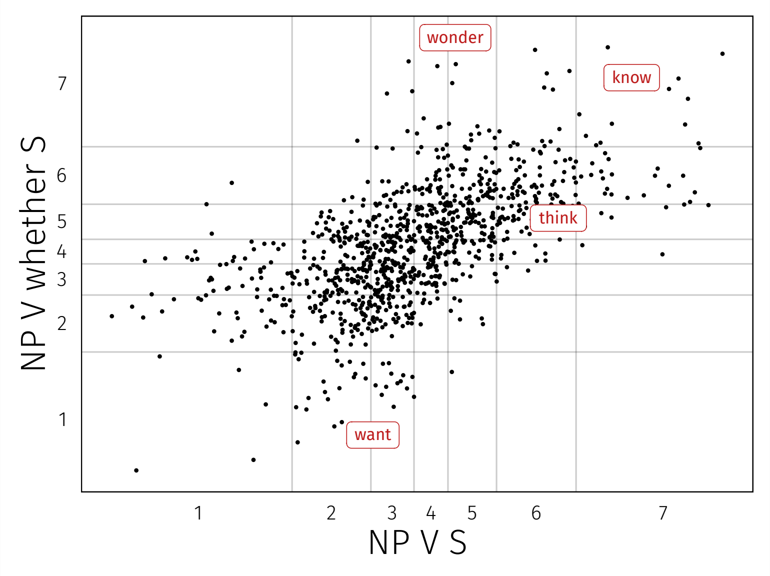


The response model

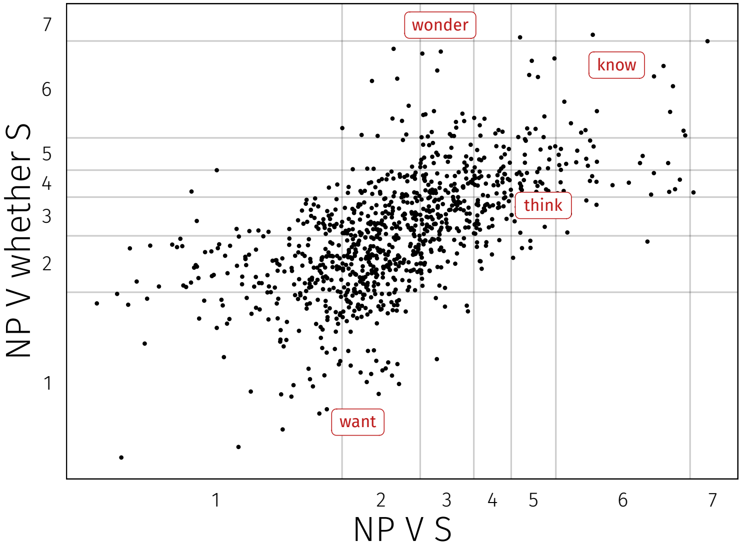



NP V whether S
7
6
5
4
3
2
1
1
2
3
4
5
6
7
NP V S

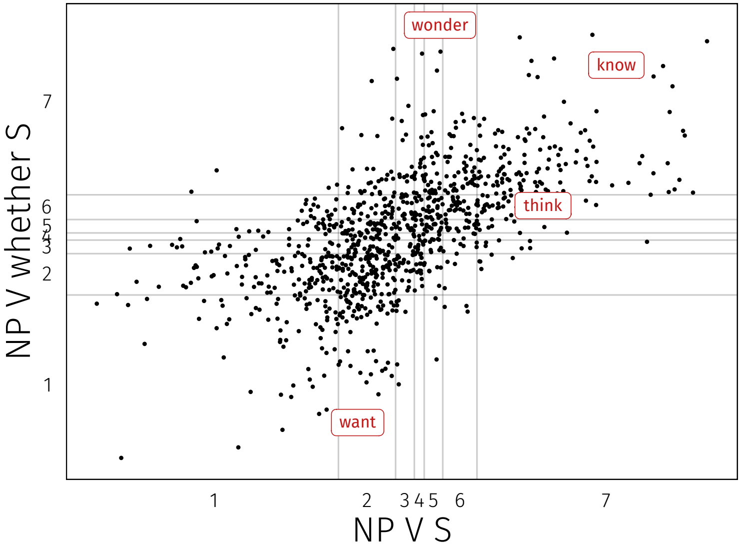
The response model



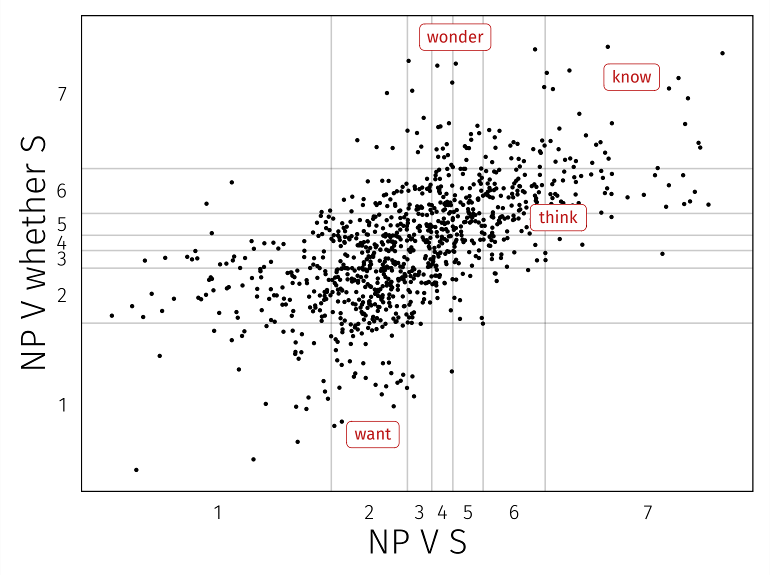
The response model



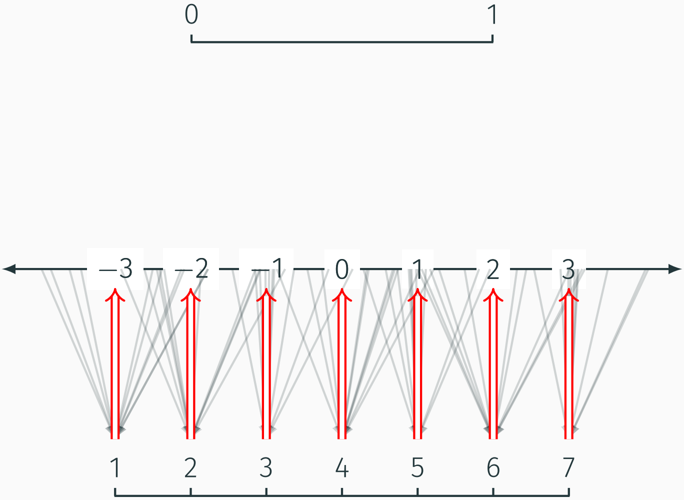
The response model



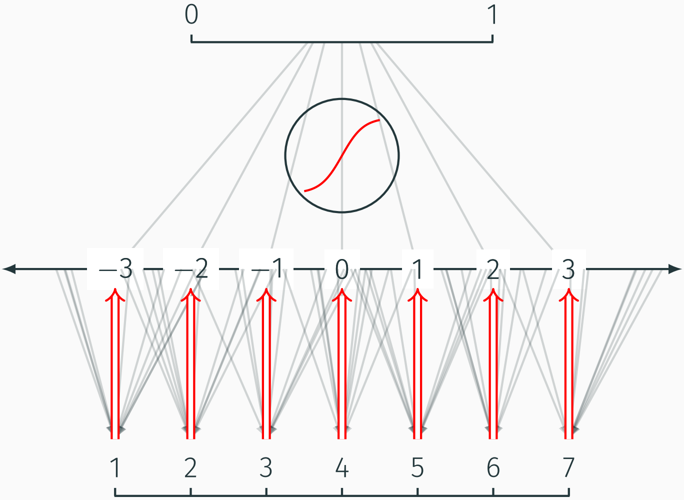
The response model



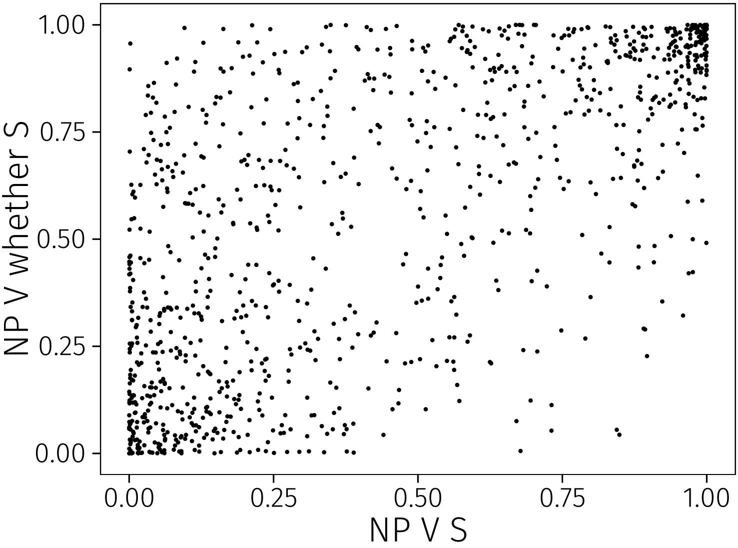
The response model



The response model



The response model



Fitting the model

Subgoal

Find the optimal number T of type signatures

Fitting the model

Subgoal

Find the optimal number T of type signatures

Goodness of $T \leftrightarrow$ model's ability to... ...fit observed judgments
...predict unobserved judgments

Fitting the model

Subgoal

Find the optimal number T of type signatures

Goodness of $T \leftrightarrow$ model's ability to... ...fit observed judgments
...predict unobserved judgments

- T too small $\rightarrow \begin{cases} \text{bad fit} \\ \text{bad prediction} \end{cases}$

Fitting the model

Subgoal

Find the optimal number T of type signatures

Goodness of $T \leftrightarrow$ model's ability to... ...fit observed judgments
...predict unobserved judgments

- T too small $\rightarrow \begin{cases} \text{bad fit} \\ \text{bad prediction} \end{cases}$
- T too large $\rightarrow \begin{cases} \text{good fit} \\ \text{bad prediction} \end{cases}$

Fitting the model

Number of type signatures

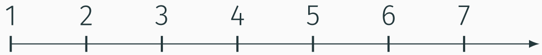
Fitting the model

Number of type signatures

Low extreme
All verbs' syntactic distributions explained by single rule

Fitting the model

Number of type signatures



Low extreme
All verbs' syntactic distributions explained by single rule

High extreme
types $\geq$ # frames
every syntactic frame has separate rule

Fitting the model

Subgoal

Find the optimal number T of type signatures

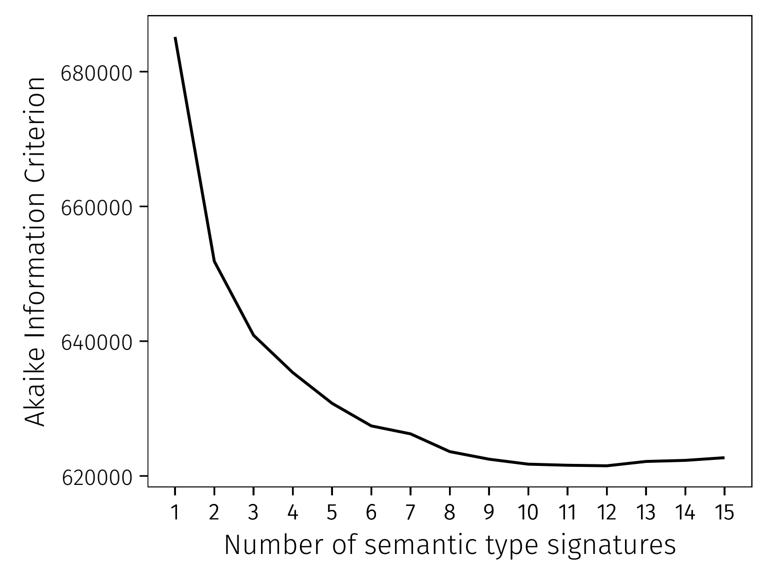
Goodness of $T \leftrightarrow$ model's ability to... ...fit observed judgments
...predict unobserved judgments

- T too small $\rightarrow \begin{cases} \text{bad fit} \\ \text{bad prediction} \end{cases}$
- T too large $\rightarrow \begin{cases} \text{good fit} \\ \text{bad prediction} \end{cases}$

Measure

Akaike Information Criterion (AIC) trades off **fit to observed data** and **prediction of unobserved data**

Model comparison



Model comparison