Computational approaches to clause selection

Aaron Steven White

University of Rochester

Department of Linguistics

Goergen Institute for Data Science

Department of Computer Science

Department of Brain & Cognitive Sciences

Selectionfest
Berlin
11th November, 2017

Slides available at **aswhite.net**

Collaborator

Kyle Rawlins

Johns Hopkins University

Department of Cognitive Science

Introduction

Three questions for a theory of selection

Structure of the domain

What types of things do predicates relate?

Three questions for a theory of selection

Structure of the domain

What types of things do predicates relate?

S(emantic)-selection

Which predicates relate which types of things?

Three questions for a theory of selection

Structure of the domain

What types of things do predicates relate?

S(emantic)-selection

Which predicates relate which types of things?

Projection rules

What is the mapping from those types to syntactic structures?

Two challenges to future progress

Main assumption

We not only have the **right architectural assumptions** for answering these questions, we have **pretty good answers**.

Main assumption

We not only have the **right architectural assumptions** for answering these questions, we have **pretty good answers**.

Two challenges

As our theories of selection gain coverage of the lexicon...

Two challenges to future progress

Main assumption

We not only have the **right architectural assumptions** for answering these questions, we have **pretty good answers**.

Two challenges

As our theories of selection gain coverage of the lexicon...

1. ...distinguishing competing theories requires more data + methods for scaling distributional analysis to those data.

Two challenges to future progress

Main assumption

We not only have the **right architectural assumptions** for answering these questions, we have **pretty good answers**.

Two challenges

As our theories of selection gain coverage of the lexicon...

1. ...distinguishing competing theories requires more data + methods for scaling distributional analysis to those data.
2. ...they grow in complexity, requiring a learning account that is capable of acquiring this complexity from a corpus.

Today's talk

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Today's talk

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Today's talk

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Basic idea

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

Today's talk

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Basic idea

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

Today's talk

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Basic idea

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

Today's talk

Focus

Syntactic distribution of $\sim$1000 English clause-embedding verbs

Today's talk

Focus

Syntactic distribution of $\sim$1000 English clause-embedding verbs

Question #1

What does the model infer about S-selection and projection, given syntactic distributions collected via acceptability judgments?

Today's talk

Focus

Syntactic distribution of ∼1000 English clause-embedding verbs

Question #1

What does the model infer about S-selection and projection, given syntactic distributions collected via acceptability judgments?

Question #2

How does the model's solution compare when given syntactic distributions collected from a corpus?

Today's talk

Idea ($\approx$ poverty of the stimulus argument)

If S-selection for some type cannot be gleaned from a corpus, an otherwise learnable semantic property determines it.

Today's talk

Idea ($\approx$ poverty of the stimulus argument)

If S-selection for some type cannot be gleaned from a corpus, an otherwise learnable semantic property determines it.

Finding

There are types that cannot be learned even from large corpora.

Today's talk

Idea ($\approx$ poverty of the stimulus argument)

If **S-selection** for some type cannot be gleaned from a corpus, an otherwise learnable **semantic property** determines it.

Finding

There are types that cannot be learned even from large corpora.

Methodological implication

We cannot rely on corpus distributions alone for determining selectional patterns.

Case study

Responsive predicates: take both interrogative and declaratives

(1) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

Today's talk

Case study

Responsive predicates: take both interrogative and declaratives

(1) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

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

Today's talk

Case study

Responsive predicates: take both interrogative and declaratives

(1) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

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

Finding #1 (based on acceptability judgments)

Different answer for communicative and cognitive verbs.

Today's talk

Case study

Responsive predicates: take both interrogative and declaratives

(1) a. John knows {that, whether} it's raining.
  b. John told Mary {that, whether} it was raining.

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

Finding #1 (based on acceptability judgments)

Different answer for communicative and cognitive verbs.

Finding #2 (based on comparison of acceptability) and corpus

Only the cognitive verb pattern is evidenced in the corpora.

Outline

A model of S-selection & projection

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 $\textbf{type}([\![NP]\!]) = \textbf{type}([\![S]\!])$

cf. Baker 1968, Heim 1979, Romero 2005, Nathan 2006, Frana 2010a, Aloni & Roelofsen 2011

Projection

Semantic type [___Q] (Grimshaw's notation)

Syntactic type [___S] [___NP]

Projection

Semantic type [___Q] (Grimshaw's notation)

Projection

Syntactic type [___S] [___NP]

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?

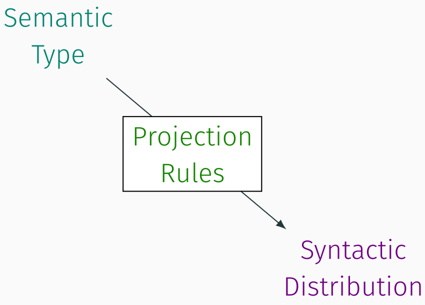
How are a verb's semantic type signatures projected onto its syntactic type signatures (subcategorization frames)? (Gruber 1965, Jackendoff 1972, Carter 1976, Grimshaw 1979, 1990, Chomsky 1981, Pesetsky 1982, 1991, Pinker 1984, 1989, Levin 1993)

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 2010b, a.o.)

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

Two kinds of lexical idiosyncrasy

The additive approach (Grimshaw 1979)

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

$\boxed{\text{S-selection} \circ \text{projection} \vee \text{C-selection} = \text{syntactic distribution}}$

Two kinds of lexical idiosyncrasy

The additive approach (Grimshaw 1979)

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

$\boxed{\text{S-selection} \circ \text{projection} \vee \text{C-selection} = \text{syntactic distribution}}$

The multiplicative approach (Pesetsky 1982, 1991)

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

$\boxed{\text{S-selection} \circ \text{projection} \wedge \text{case} = \text{syntactic distribution}}$

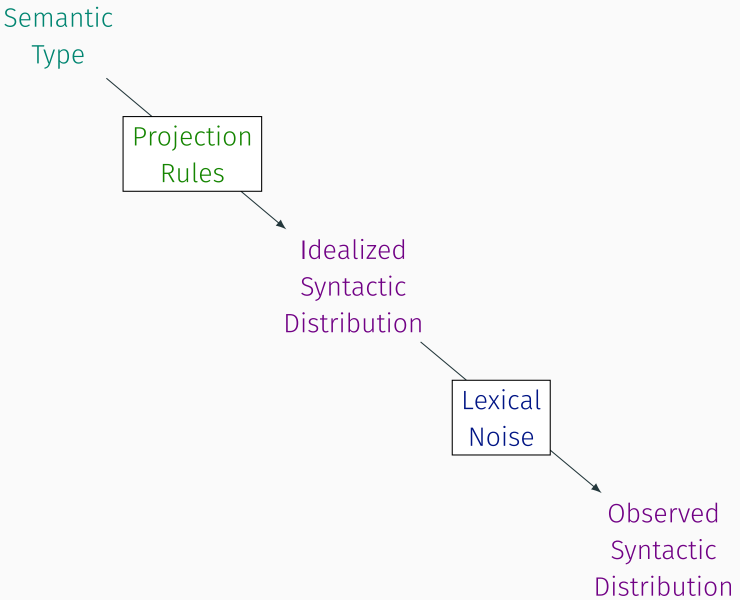
Two kinds of lexical idiosyncrasy

Shared core see White & Rawlins 2016 for formal details

Lexical noise—i.e. lexical idiosyncrasy—alters idealized syntactic distributions

$\boxed{\text{S-selection} \circ \text{projection} \otimes \text{noise} = \text{syntactic distribution}}$

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

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?

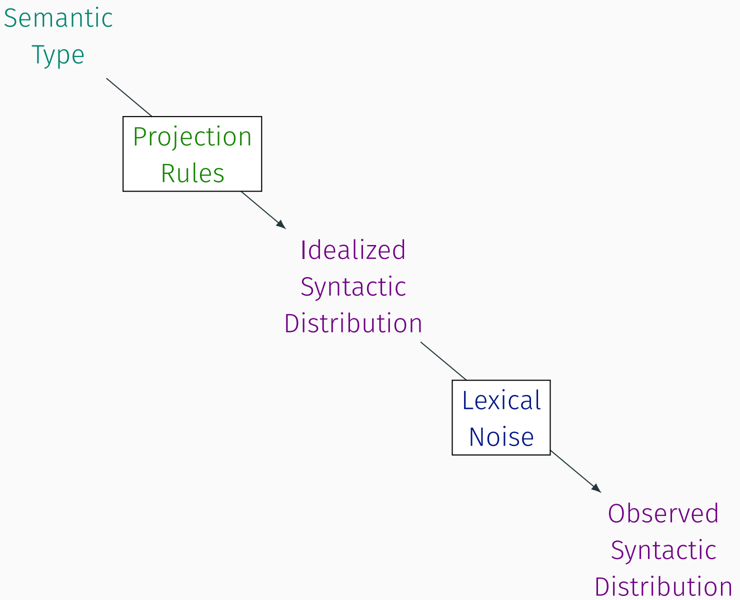
A minimalistic answer

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

know $\rightarrow$ {[___P], [___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

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

$$\mathsf{S} = \begin{array}{ll} & \begin{array}{ccc} [\_\_\_\text{P}] & [\_\_\_\text{Q}] & \cdots \end{array} \\ \begin{array}{l} \text{think} \\ \text{know} \\ \text{wonder} \\ \ldots \end{array} & \begin{pmatrix} 1 & 0 & \cdots \\ 1 & 1 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \end{array}$$

A boolean model of projection

$[\_\_\_\text{P}] \rightarrow \{[\_\_\_\text{that S}], [\_\_\_\text{NP}], \ldots\}$ $\qquad$ $[\_\_\_\text{Q}] \rightarrow \{[\_\_\_\text{whether S}], [\_\_\_\text{NP}], \ldots\}$

A boolean model of projection

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

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

A boolean model of idealized syntactic distribution

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

A boolean model of idealized syntactic distribution

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

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

A boolean model of idealized syntactic distribution

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

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

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

A boolean model of idealized syntactic distribution

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

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

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

A boolean model of idealized syntactic distribution

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

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

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

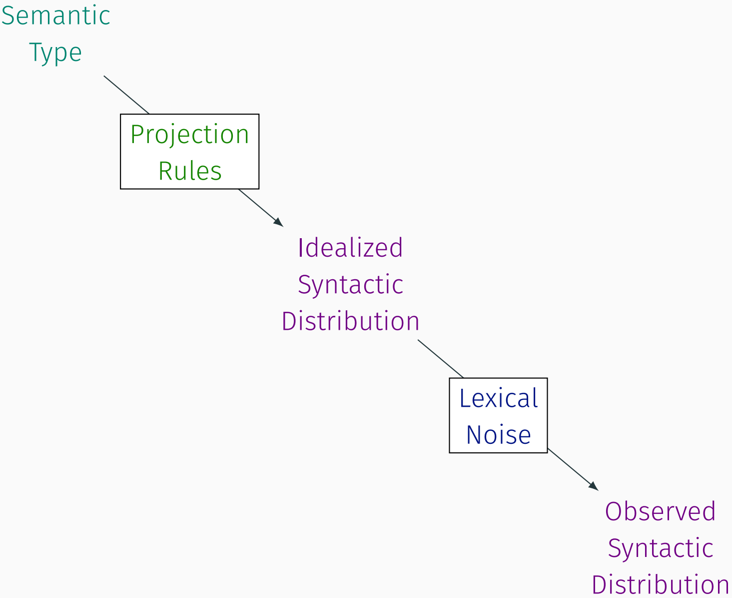
A boolean model of idealized syntactic distribution

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

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

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

A model of S-selection and projection



A boolean model of observed syntactic distribution

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

A boolean model of observed syntactic distribution

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

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

A boolean model of observed syntactic distribution

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

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

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

A boolean model of observed syntactic distribution

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

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

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

Question

What is this model useful for?

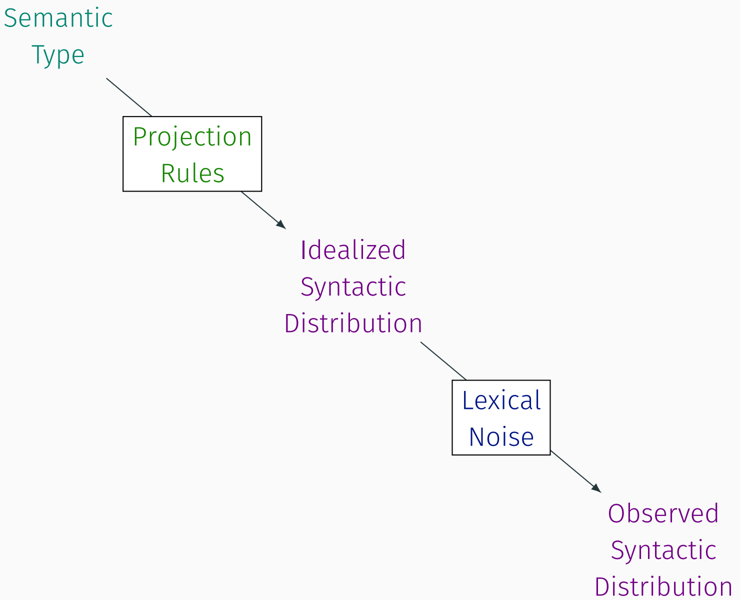
Answer

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

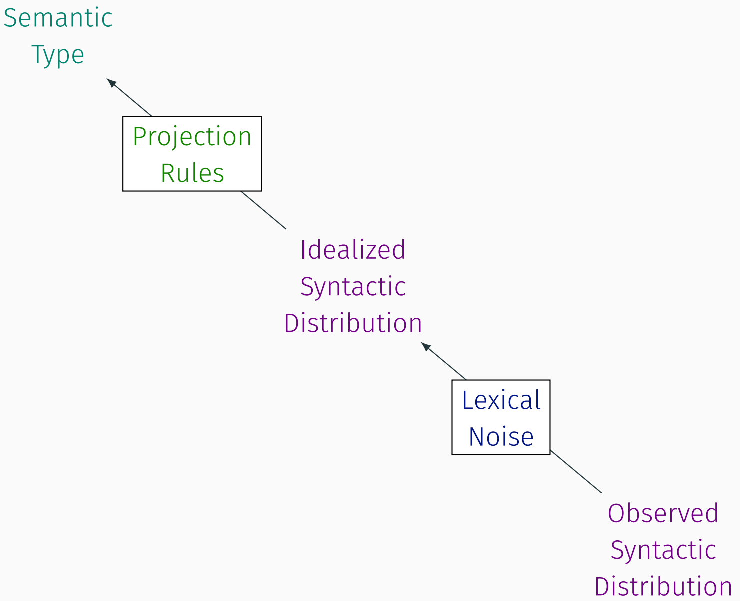
Basic idea

Distributional analysis corresponds to reversing model arrows

A model of S-selection and projection



A model of S-selection and projection



Acceptability dataset

Data available at **megaattitude.com**

MegaAttitude materials

Ordinal (1-7 scale) acceptability ratings

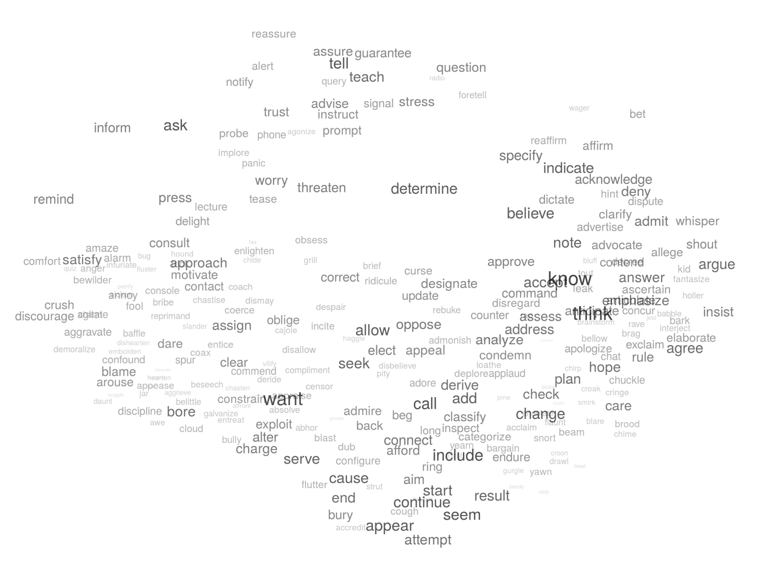
MegaAttitude materials

Ordinal (1-7 scale) acceptability ratings

for

1000 clause-embedding verbs

Verb selection



MegaAttitude materials

Ordinal (1-7 scale) acceptability ratings

for

1000 clause-embedding verbs

$\times$

50 syntactic frames

Challenge

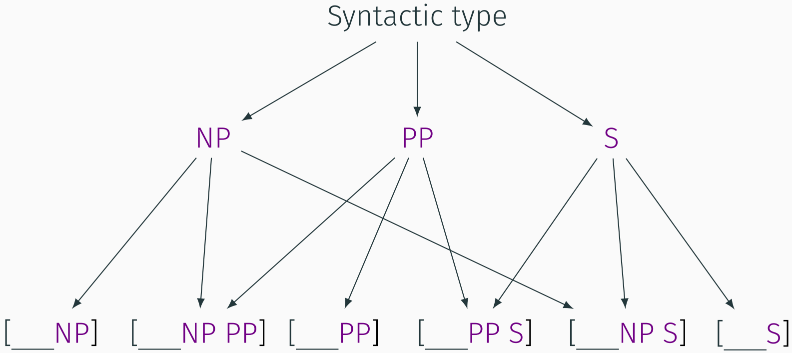
Automate construction of a very large set of frames in a way that is sufficiently general to many verbs

Frame construction

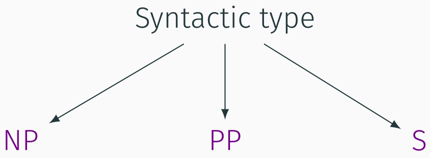
Syntactic type

NP PP S

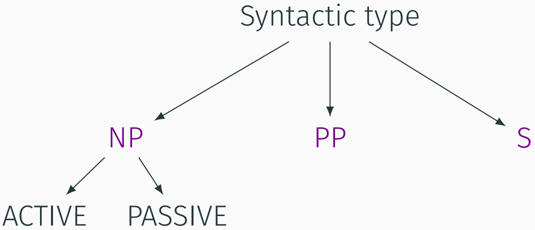
Frame construction



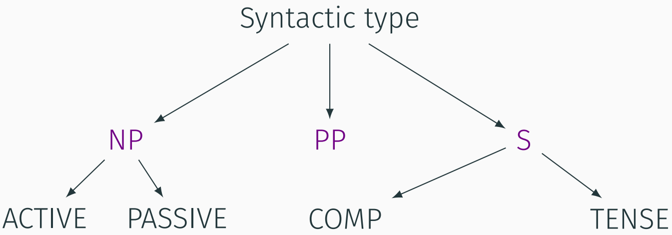
Frame construction

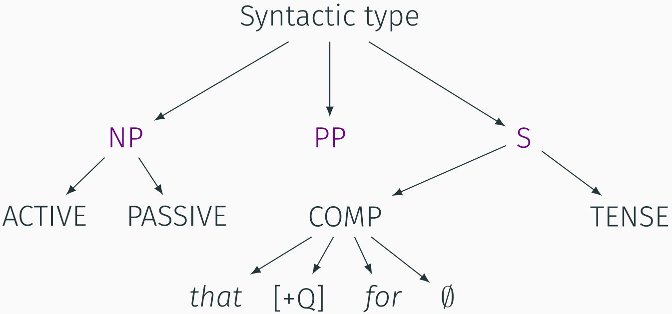


Frame construction

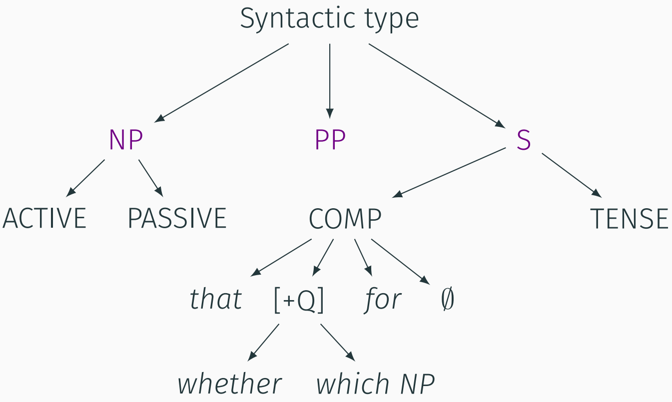


Frame construction

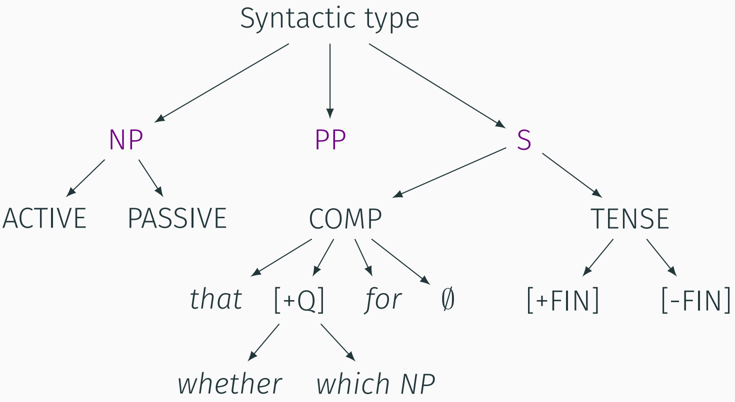



Syntactic type
NP
PP
S
ACTIVE
PASSIVE
COMP
TENSE
that
[+Q]
for
∅

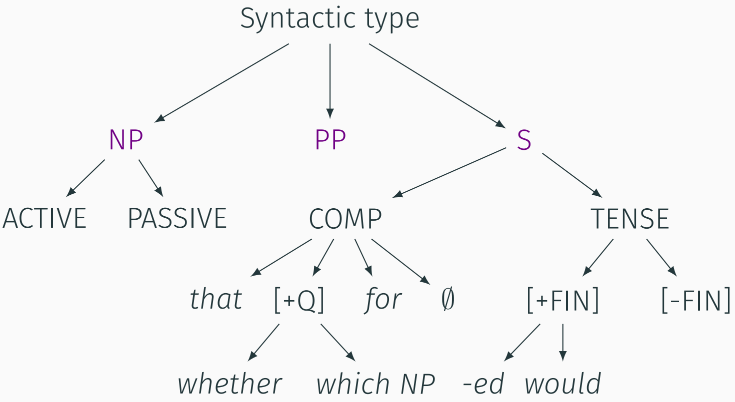
Frame construction



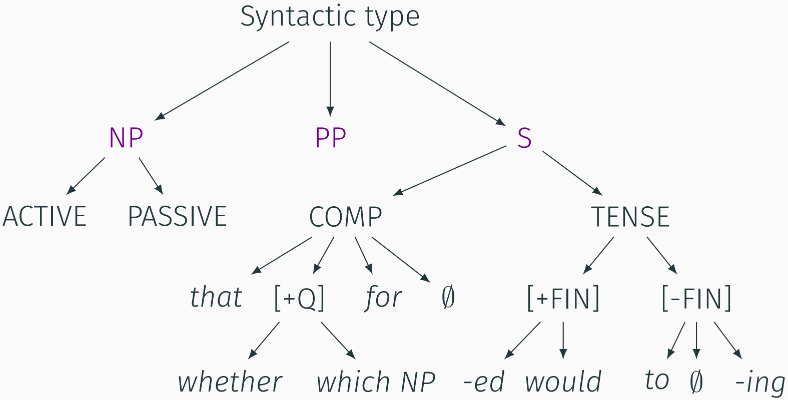
Frame construction



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

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.

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.

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

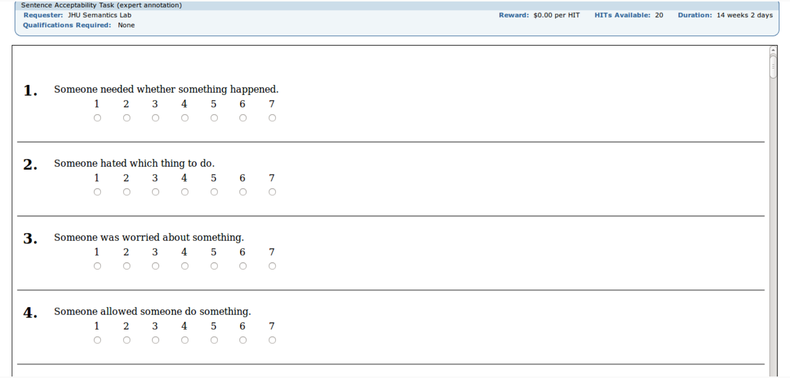
Data collection

- 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

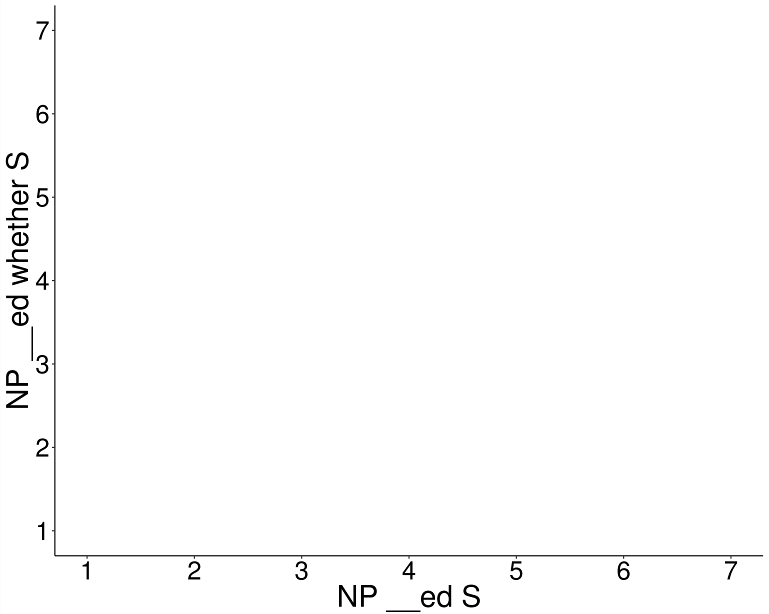
Spearman rank correlation calculated by list on a pilot 30 verbs

Pilot verb selection

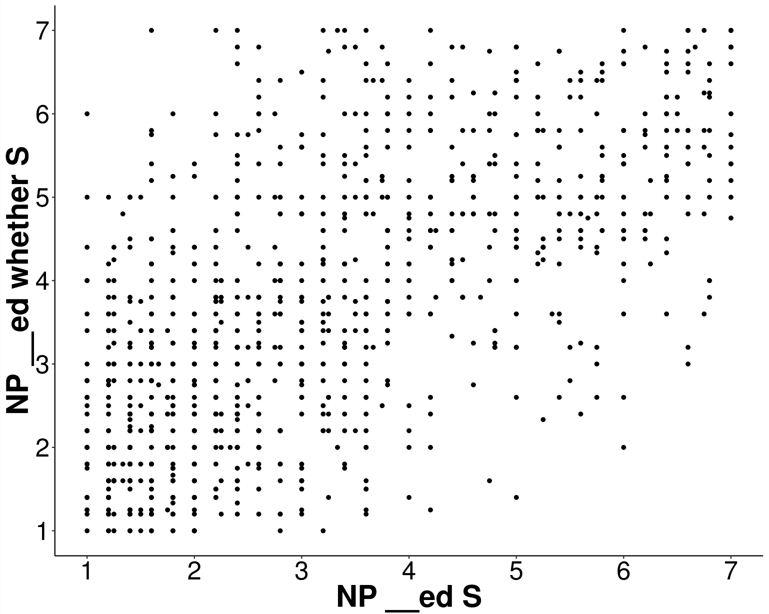
Same verbs used by White (2015), White et al. (2015), selected based on Hacquard & Wellwood's (2012) attitude verb classification

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]

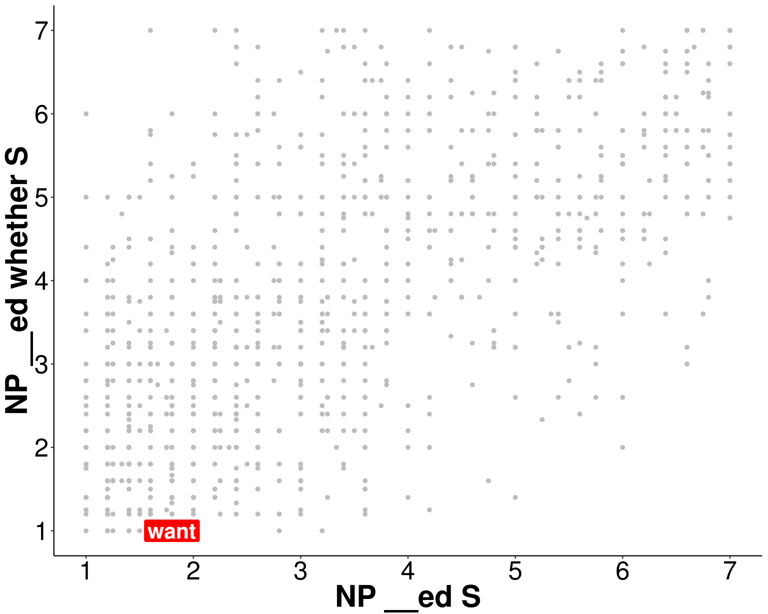
Results



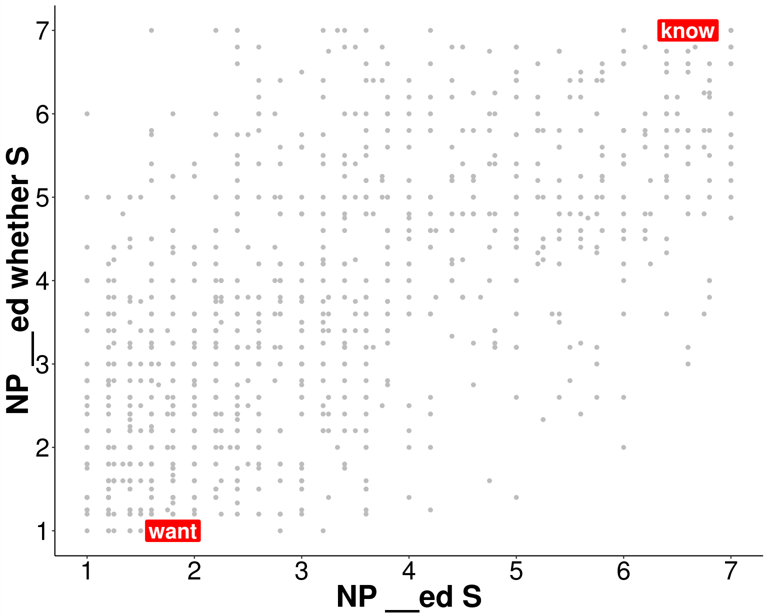
Results



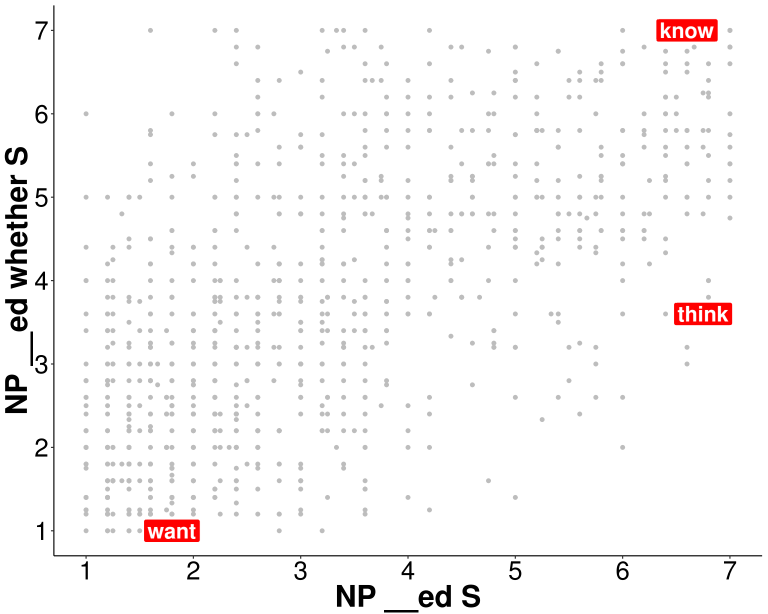
Results



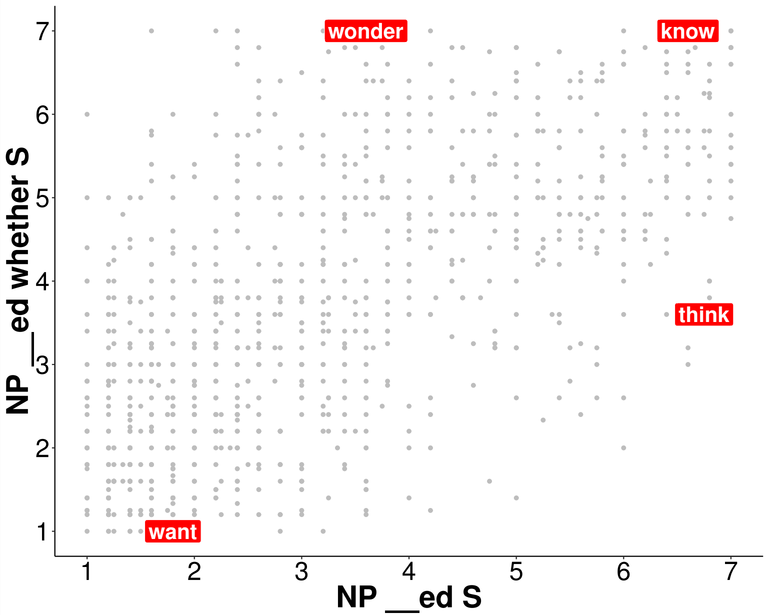
Results



Results

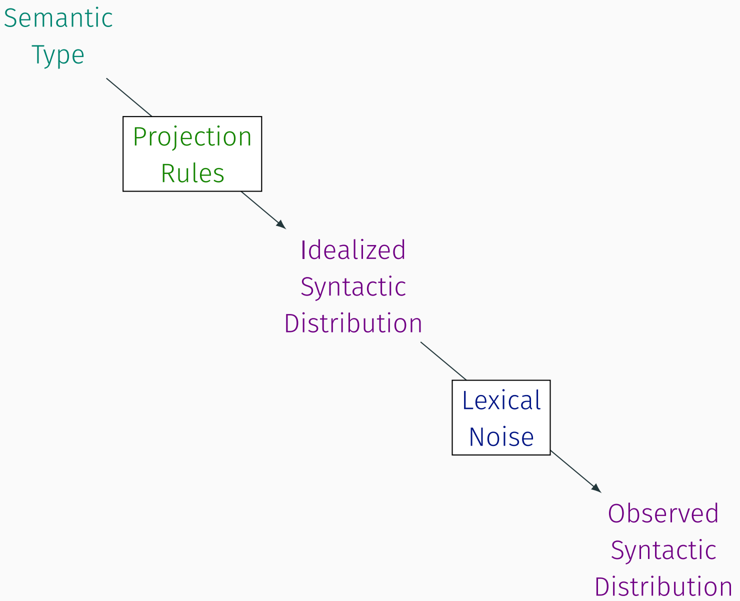


Results

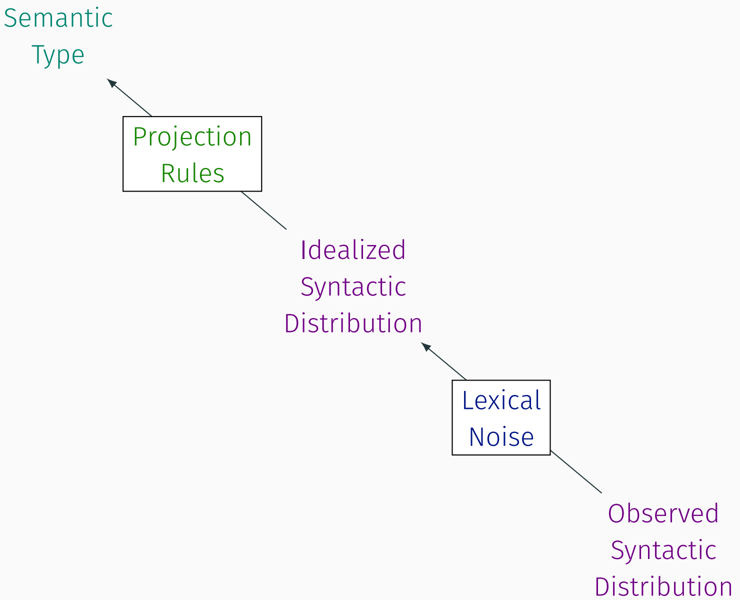


wonder
know
think
want
7
6
5
4
3
2
1
1
2
3
4
5
6
7
NP __ed whether S
NP __ed S

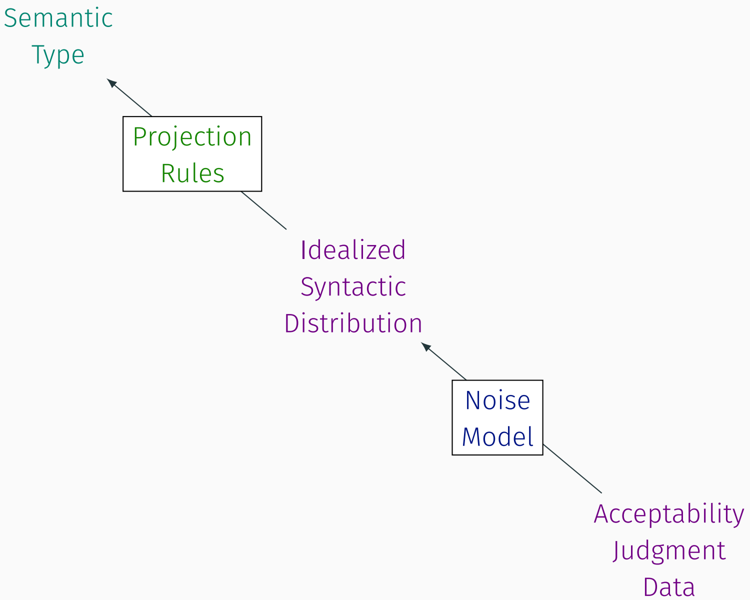
A model of S-selection and projection



A model of S-selection and projection



A model of S-selection and projection



Fitting the model

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)

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

Fitting the model

Solution

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

Fitting the model

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

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

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

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

A boolean model of idealized syntactic distribution

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

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

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

Wrapping with probabilities

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

Fitting the model

Noise model

Standard model for acceptability judgments: cumulative link logit mixed effects model (Agresti 2014)

Fitting the model

Noise model

Standard model for acceptability judgments: cumulative link logit mixed effects model (Agresti 2014)

Algorithm

Adam optimizer (basically, fancy gradient descent) (Kingma & Ba 2014)

Fitting the model

Noise model

Standard model for acceptability judgments: cumulative link logit mixed effects model (Agresti 2014)

Algorithm

Adam optimizer (basically, fancy gradient descent) (Kingma & Ba 2014)

Remaining challenge

Don't know the number of type signatures T

Fitting the model

Noise model

Standard model for acceptability judgments: cumulative link logit mixed effects model (Agresti 2014)

Algorithm

Adam optimizer (basically, fancy gradient descent) (Kingma & Ba 2014)

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)

Result

12 is the optimal number of type signatures according to 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)

Result

12 is the optimal number of type signatures according to AIC

Reporting findings

Best model with 12 type signatures

Findings

Three findings

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

Findings

[___P] [___Q]

Findings

$$\begin{tikzcd} {[\_\_\_ P]} \arrow[d] & {[\_\_\_ Q]} \arrow[d] \\ {[\_\_\_ \text{that } S]} & {[\_\_\_ \text{whether } S]} \end{tikzcd}$$

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

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

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]
2. Communicative predicates
   2.1 Two unified type signatures [___(Ent) P⊕Q] (optional recipient) and [___Ent P⊕Q] (obligatory recipient)

Findings

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

$$\begin{tikzcd}
& {[___\,\text{(Ent) } P \oplus Q]} \arrow[dl] \arrow[dr] & \\
{[___\,\text{to NP that S}]} & & {[___\,\text{to NP whether S}]}
\end{tikzcd}$$

Findings

[___P] [___Q]

[___that S] [___whether S]

[___(Ent) P$\oplus$Q]

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

Hybrid types

Question

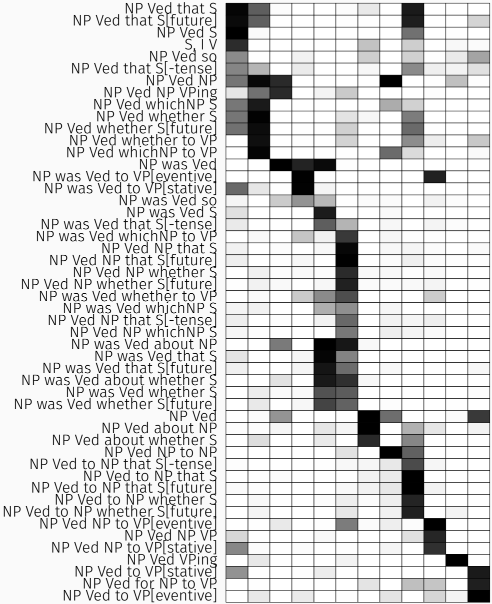
What do I mean by $P \oplus Q$?

Example

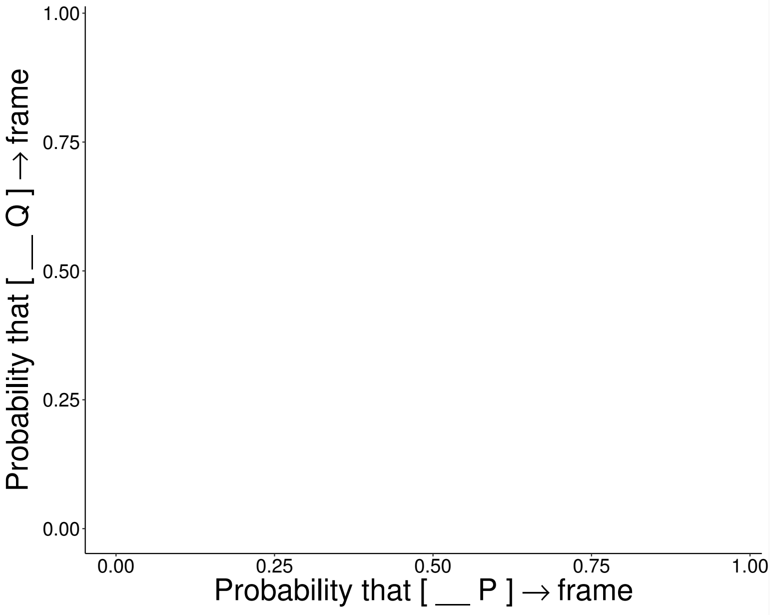
Structures with both informative and inquisitive content (Groenendijk & Roelofsen 2009, a.o.)

- 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*)

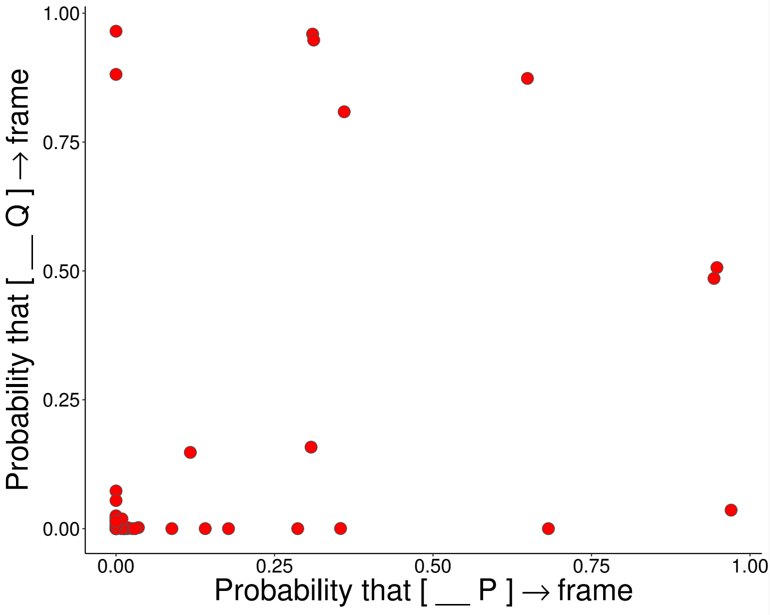
Projection


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]

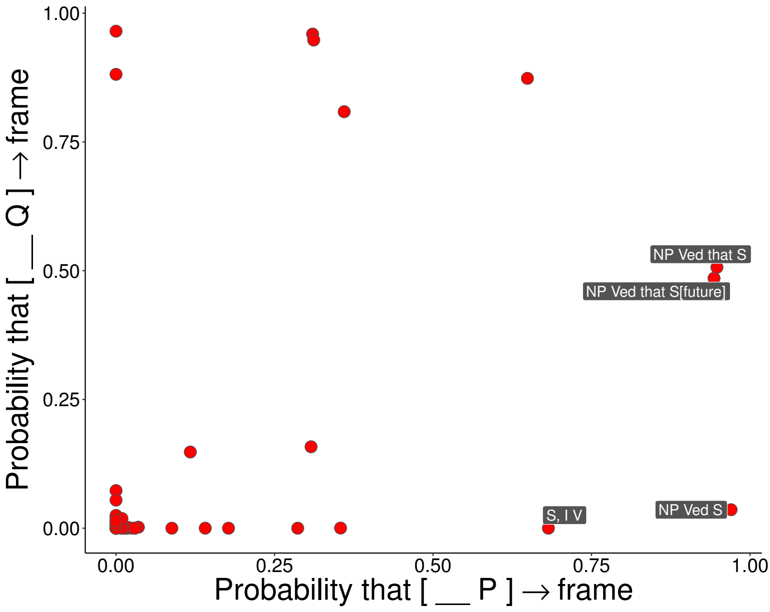
Projection: propositions and questions



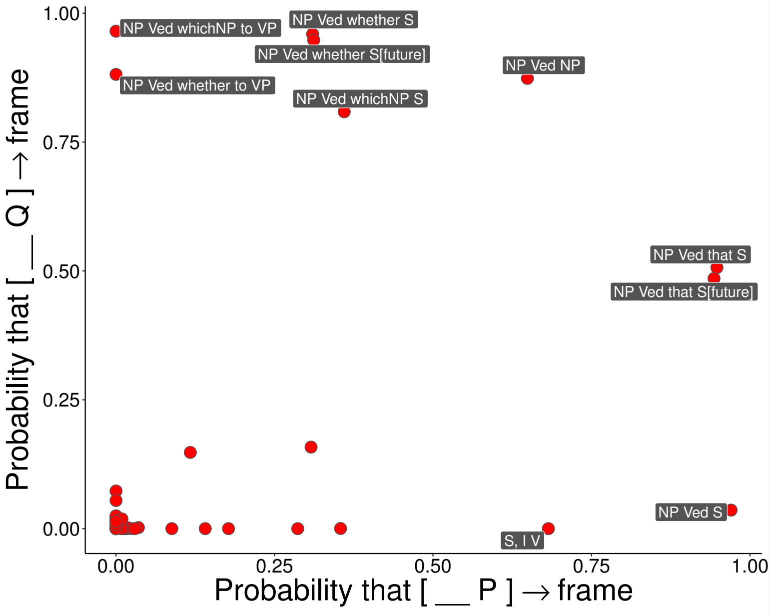
Projection: propositions and questions



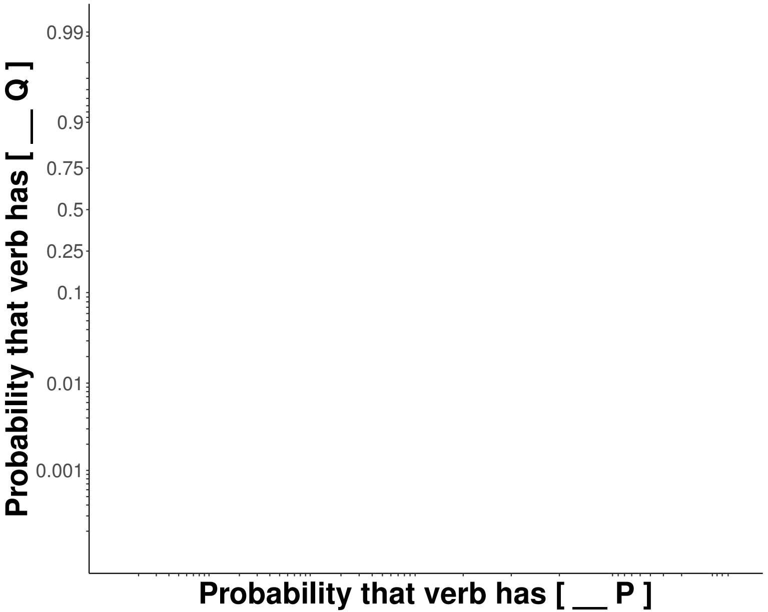
Projection: propositions and questions



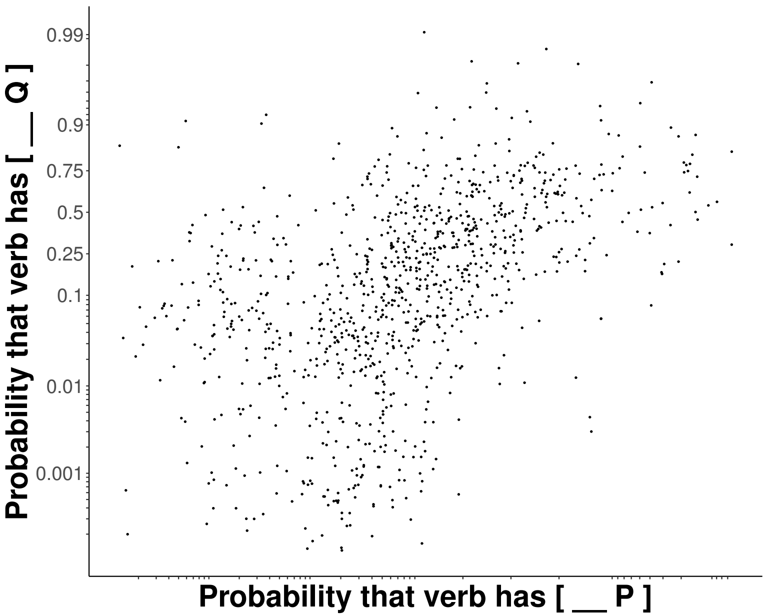
Projection: propositions and questions



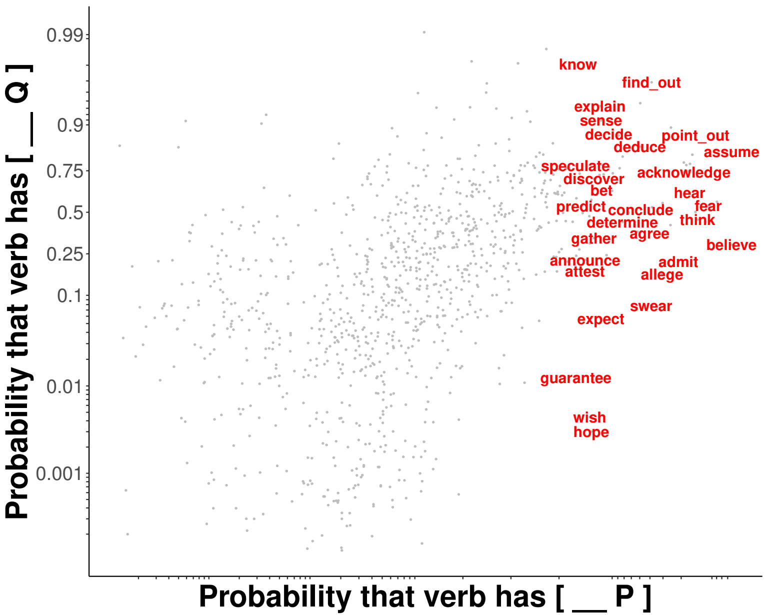
S-selection: propositions and questions



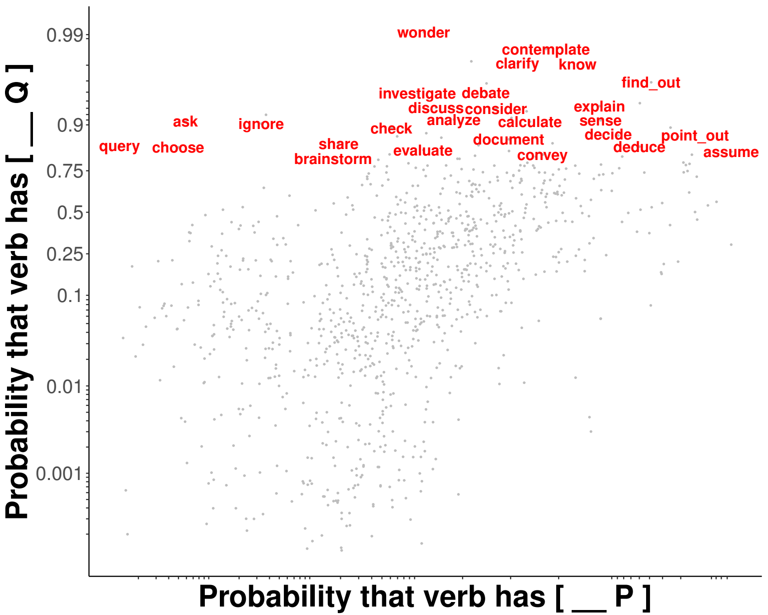
S-selection: propositions and questions



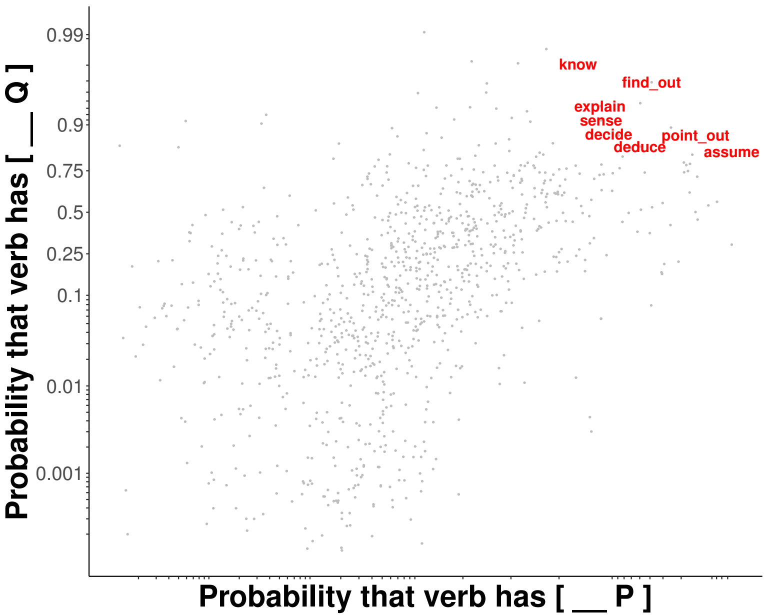
S-selection: propositions and questions



S-selection: propositions and questions



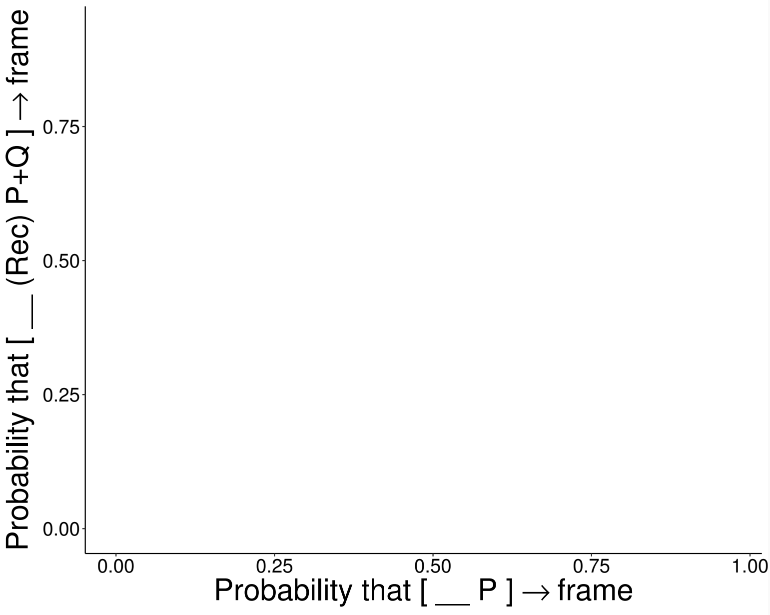
S-selection: propositions and questions



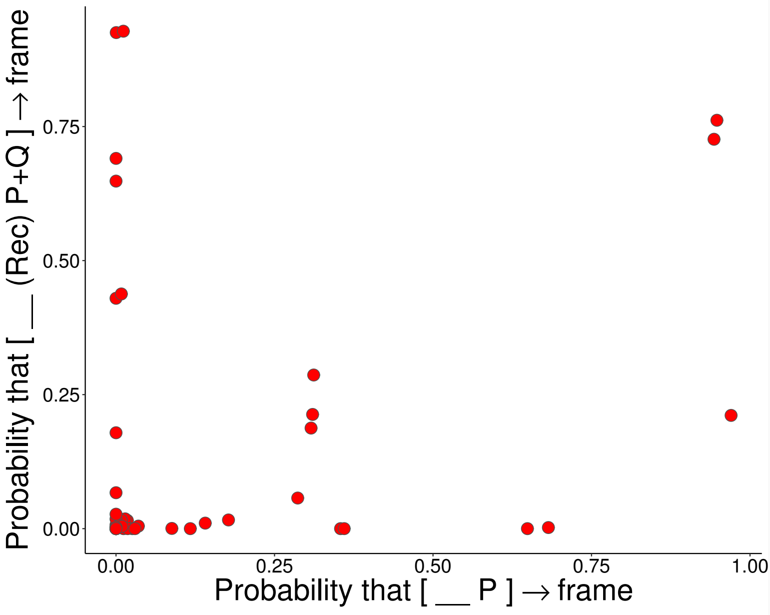
Findings

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

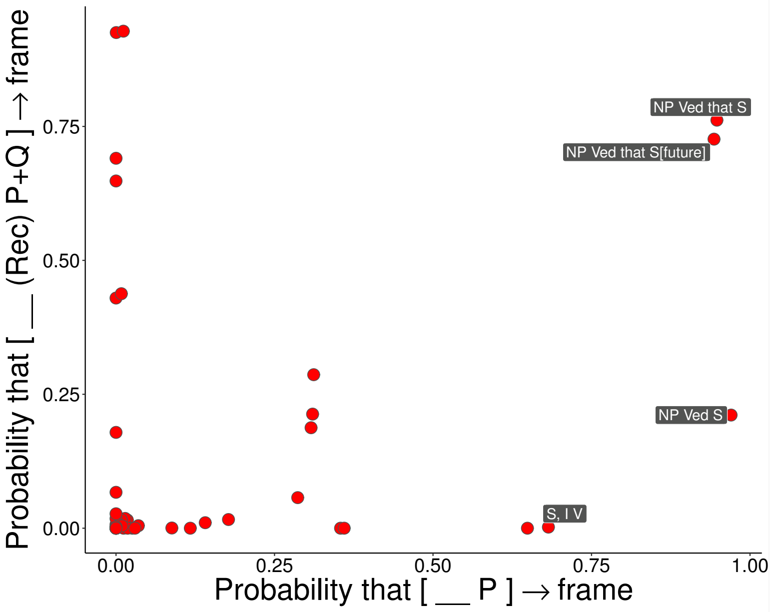
Projection: optional recipients



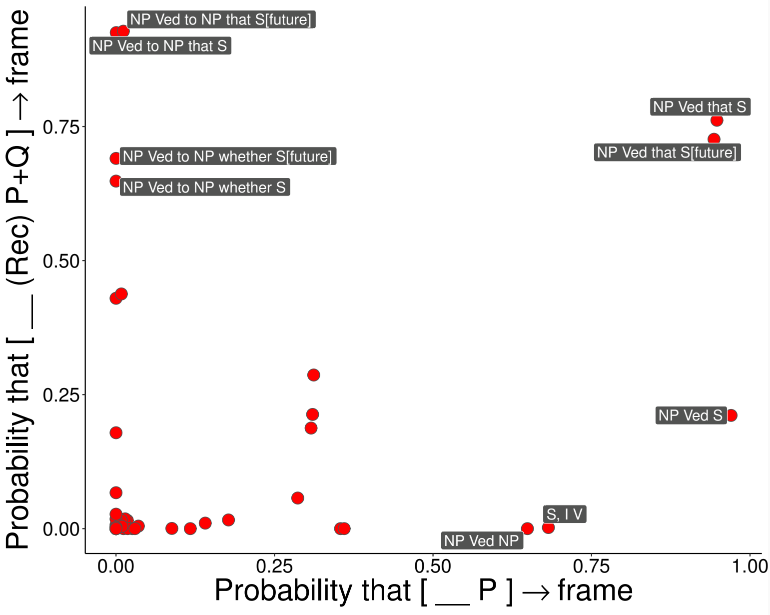
Projection: optional recipients



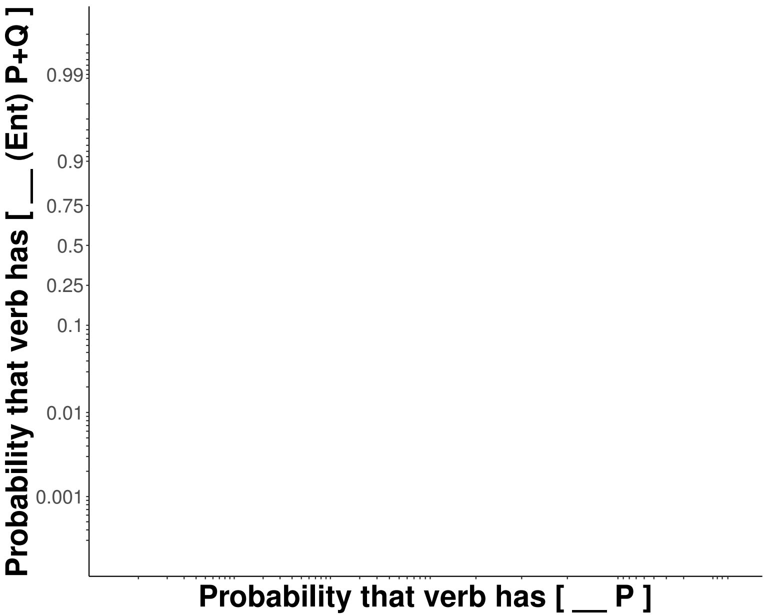
Projection: optional recipients



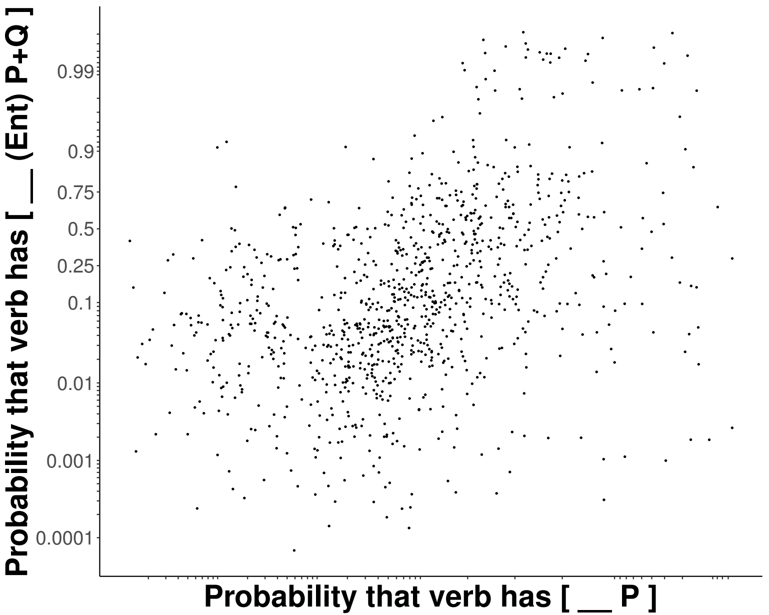
Projection: optional recipients



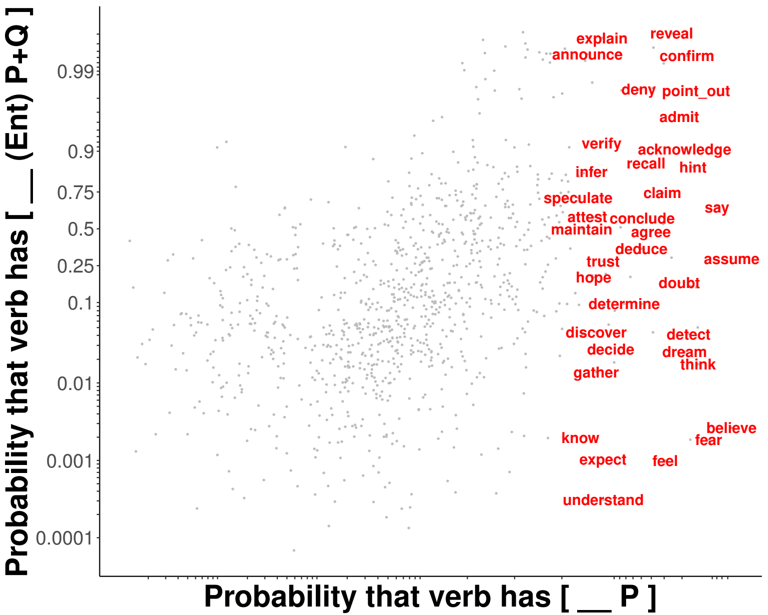
S-selection: optional recipients



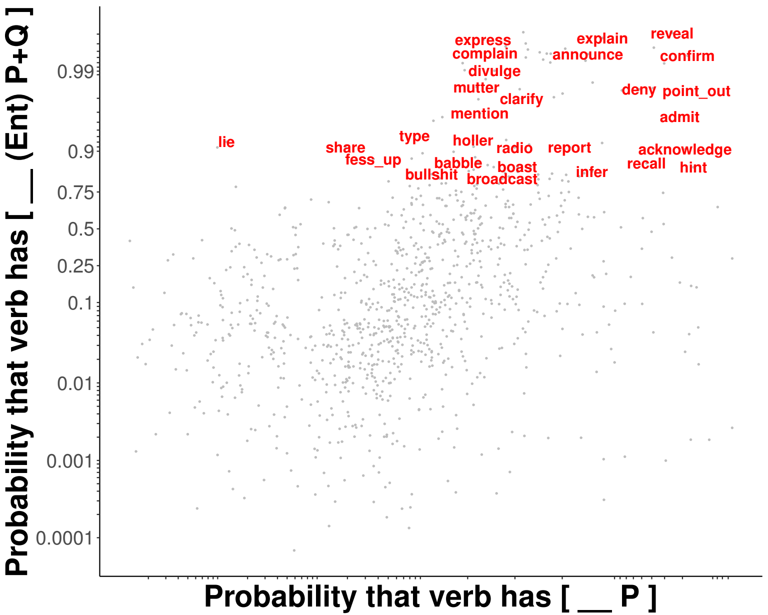
S-selection: optional recipients



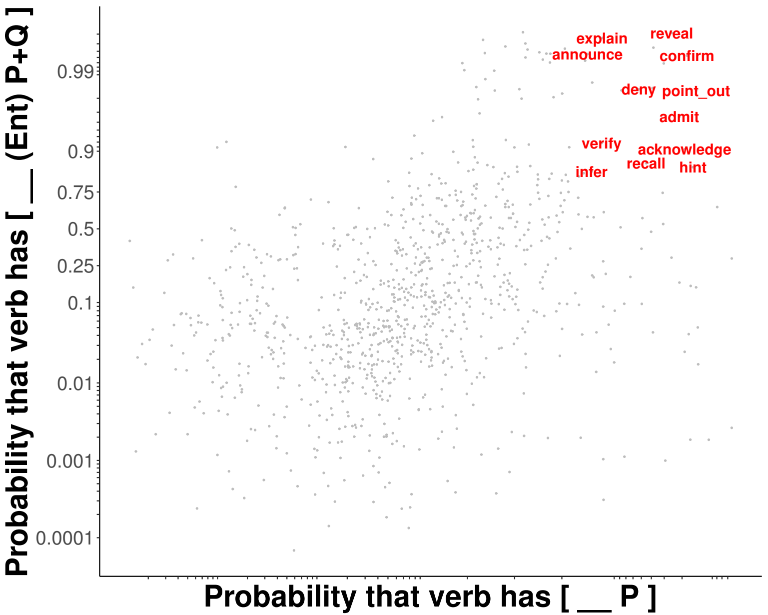
S-selection: optional recipients



S-selection: optional recipients



S-selection: optional recipients



Findings

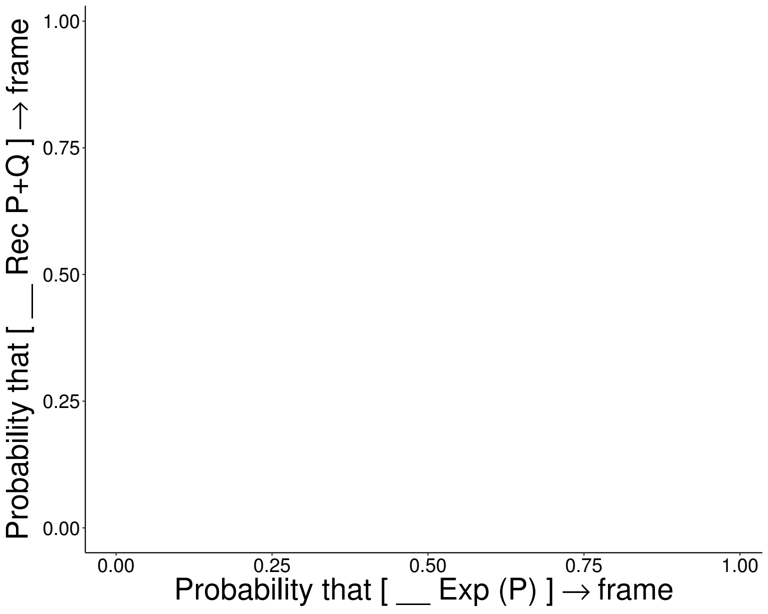
[___P] [___Q]

[___that S] [___whether S]

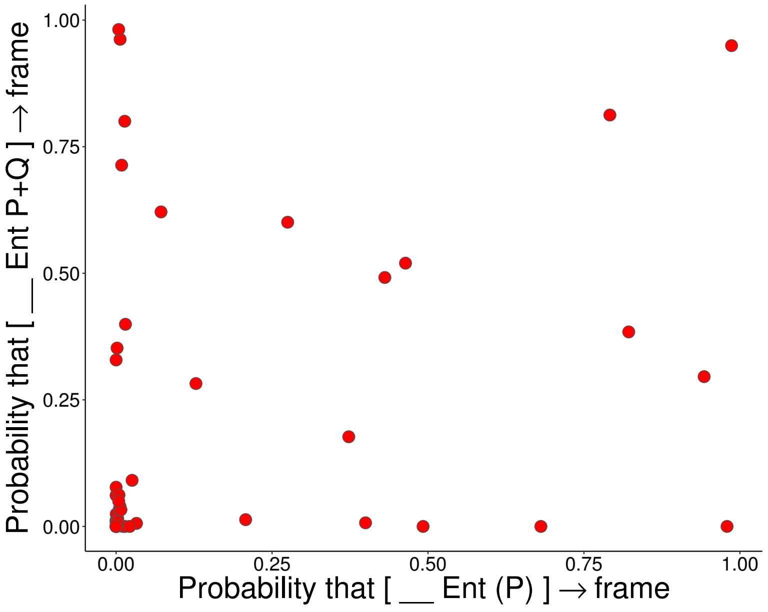
[___(Ent) P⊕Q]

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

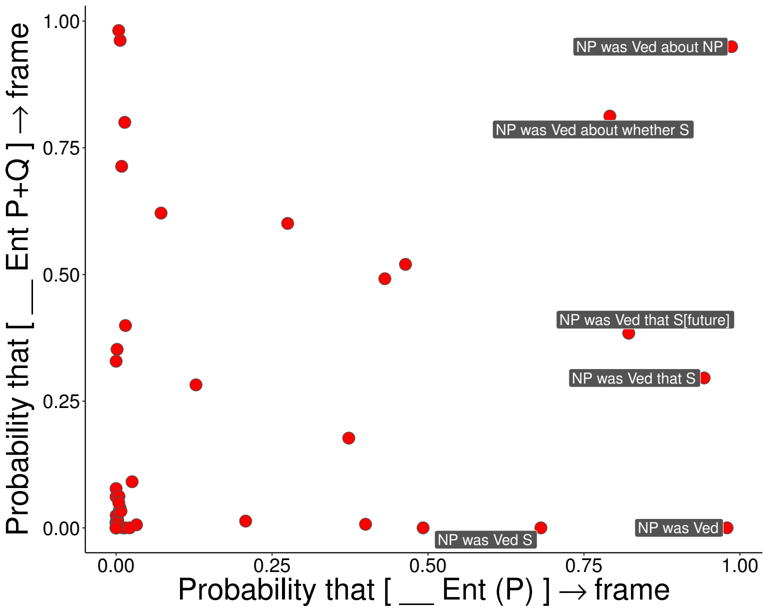
Projection: obligatory recipients/experiencers



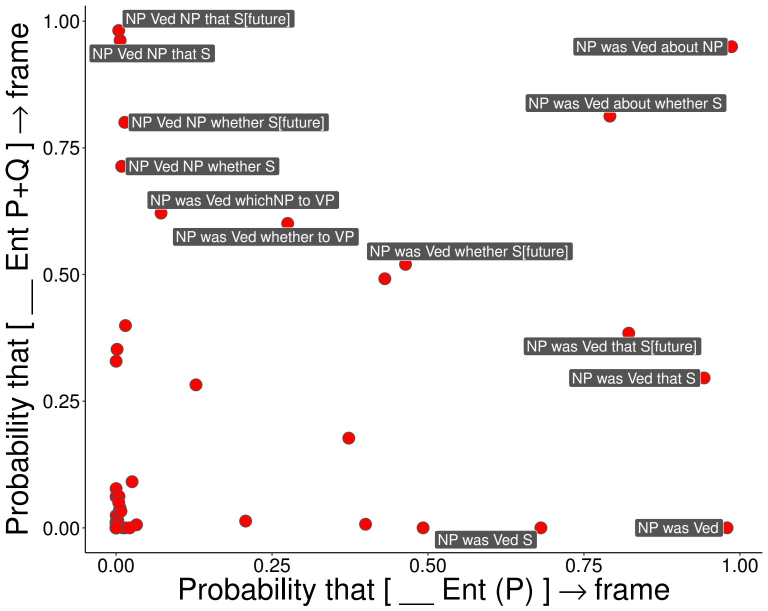
Projection: obligatory recipients/experiencers



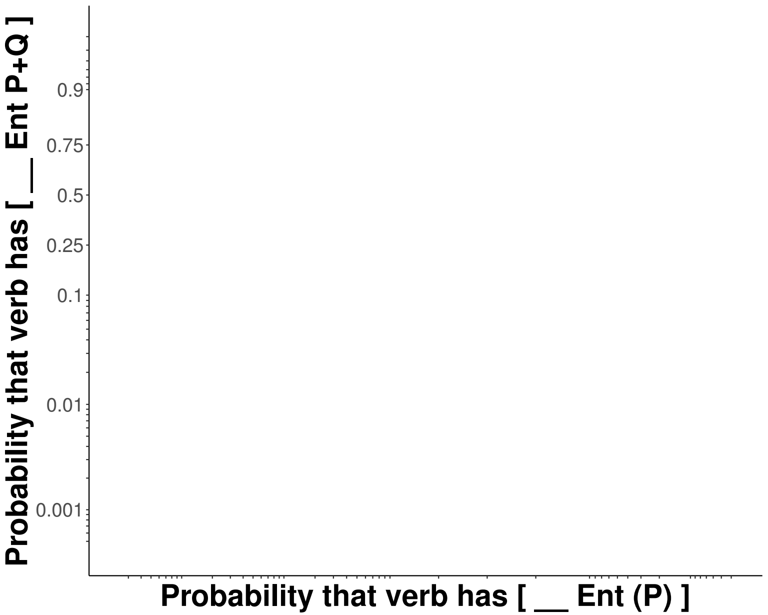
Projection: obligatory recipients/experiencers



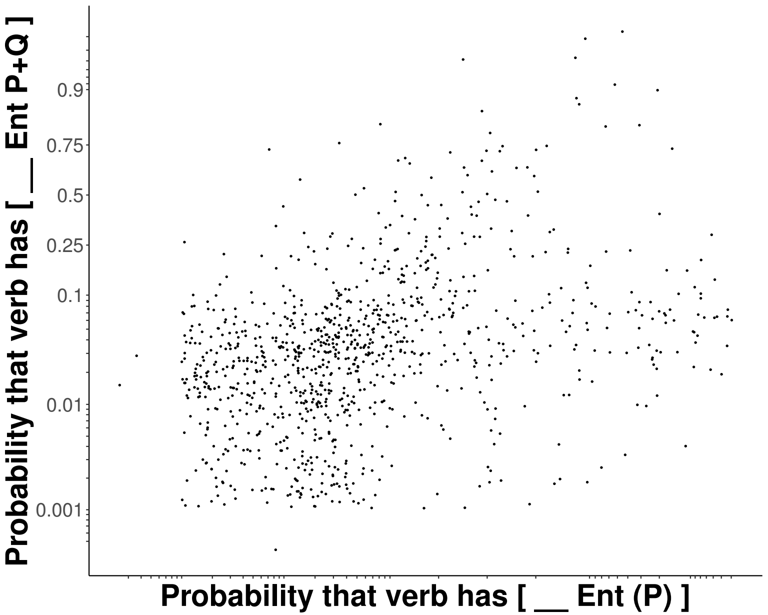
Projection: obligatory recipients/experiencers



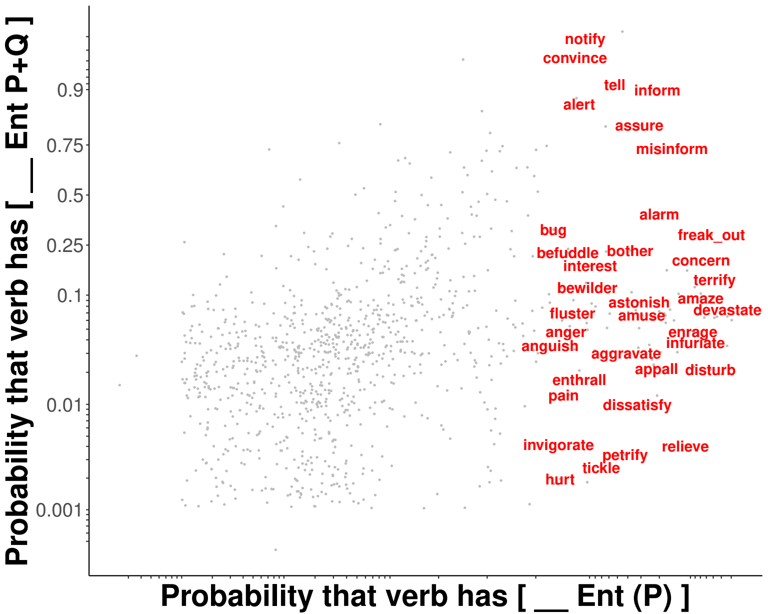
S-selection: obligatory recipients/experiencers



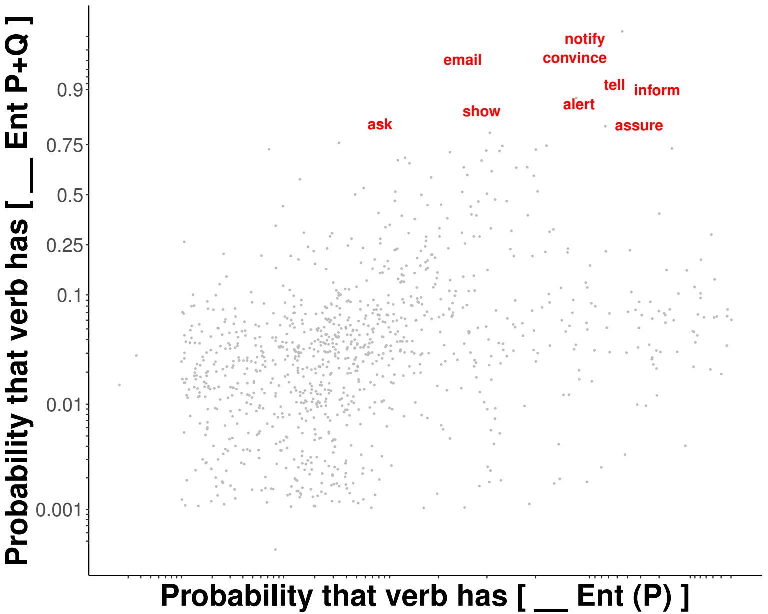
S-selection: obligatory recipients/experiencers



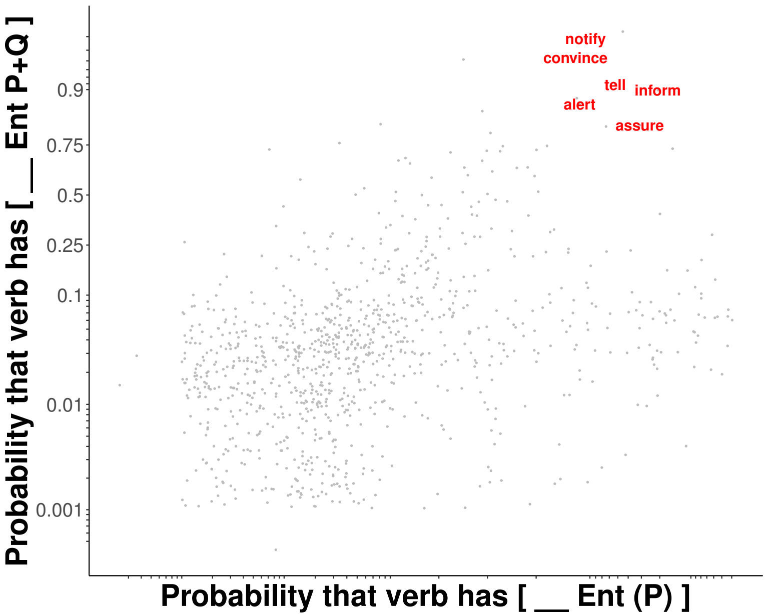
S-selection: obligatory recipients/experiencers



S-selection: obligatory recipients/experiencers



S-selection: obligatory recipients/experiencers



Findings

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

$$
\begin{tikzcd}
 & {[\_\_\_ \text{Ent P} \oplus \text{Q}]} \arrow[dl] \arrow[dr] & \\
{[\_\_\_ \text{NP that S}]} & & {[\_\_\_ \text{NP whether S}]}
\end{tikzcd}
$$

What to conclude

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

Interim discussion

What to conclude

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

What to exclude

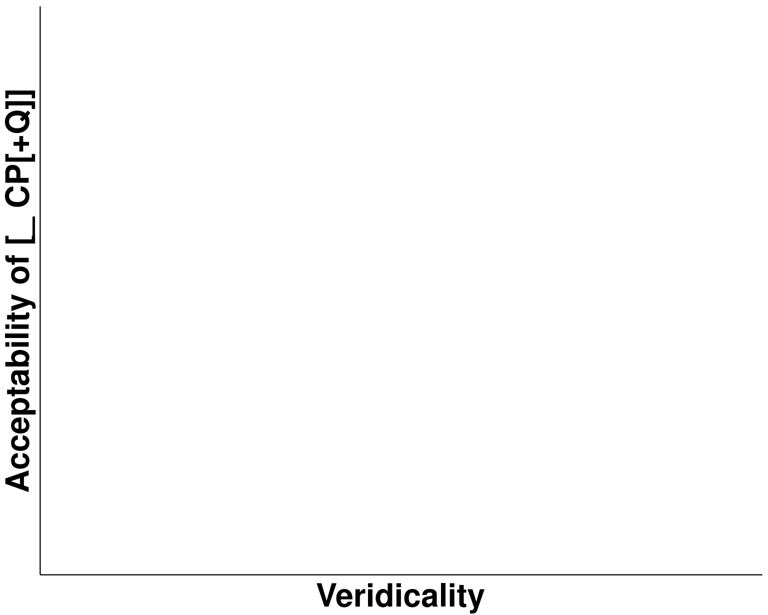
Accounts that reduce (or unify) declarative and interrogative selection solely to S-selection of a single type + coercion

Interim discussion

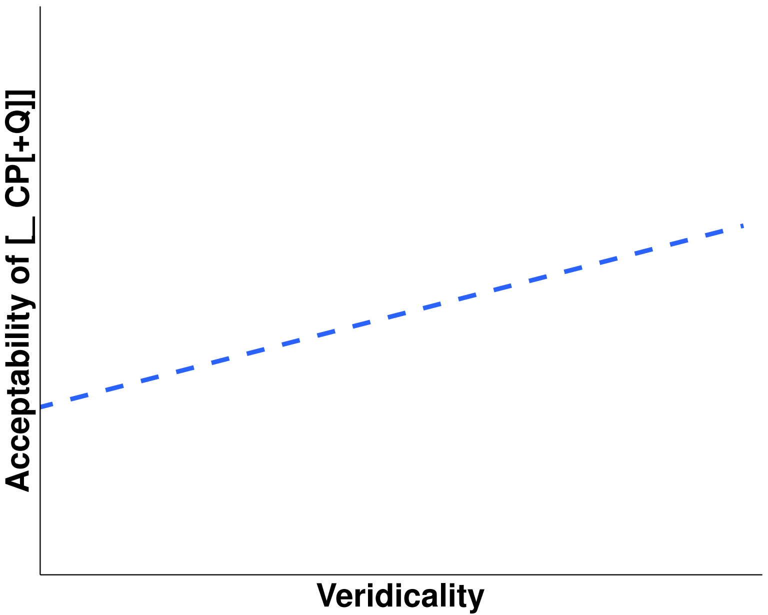
Question

Is there anything to say about whether selection for P, Q, or P⊕Q is reducible to lexical semantics?

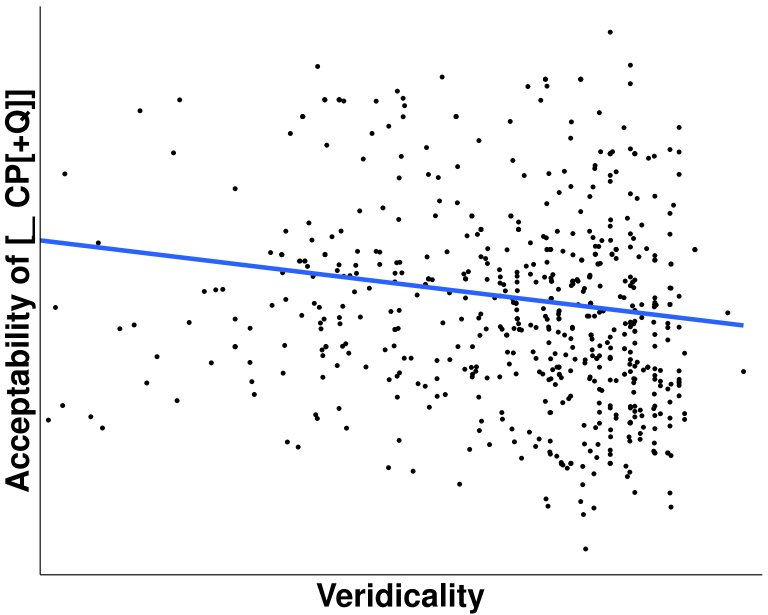
Interim discussion



Interim discussion



Interim discussion



Question

Is there anything to say about whether selection for P, Q, or $P \oplus Q$ is reducible to lexical semantics?

Interim discussion

Question

Is there anything to say about whether selection for P, Q, or P⊕Q is reducible to lexical semantics?

White & Rawlins's (2017) claim

It's all about the event structure of the predicate.

Question

Is there anything to say about whether selection for P, Q, or P⊕Q is reducible to lexical semantics?

White & Rawlins's (2017) claim

It's all about the event structure of the predicate.

Today's strategy

Do we find the same type signatures when fitting the model to corpus data?

Corpus Dataset

Corpus data

42.8 million verb-subcategorization frame pairs extracted from Parsed ukWaC (PukWaC) (Baroni et al. 2009)

Corpus data

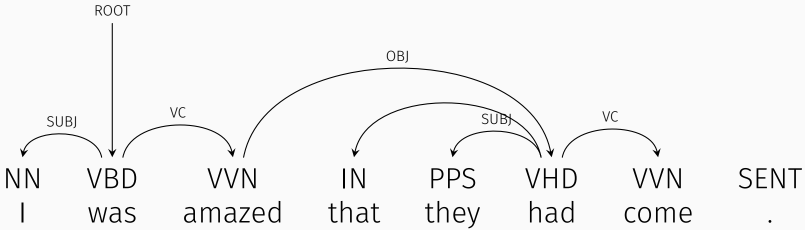
42.8 million verb-subcategorization frame pairs extracted from Parsed ukWaC (PukWaC) (Baroni et al. 2009)

2 billion word web corpus constructed from crawl of the .uk domain, dependency parsed with MaltParser (Nivre et al. 2007)

Corpus data

42.8 million verb-subcategorization frame pairs extracted from Parsed ukWaC (PukWaC) (Baroni et al. 2009)

2 billion word web corpus constructed from crawl of the .uk domain, dependency parsed with MaltParser (Nivre et al. 2007)



Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)

Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects

Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements

Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements
5. Whether there is a clausal complement, and if so...

Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements
5. Whether there is a clausal complement, and if so…
    5.1 …what the complementizer is (if any)

Subcategorization frame extraction

Features extracted see White 2015 for details

1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements
5. Whether there is a clausal complement, and if so...
   5.1 ...what the complementizer is (if any)
   5.2 ...what the WH word is (if any)

Subcategorization frame extraction

Features extracted see White 2015 for details

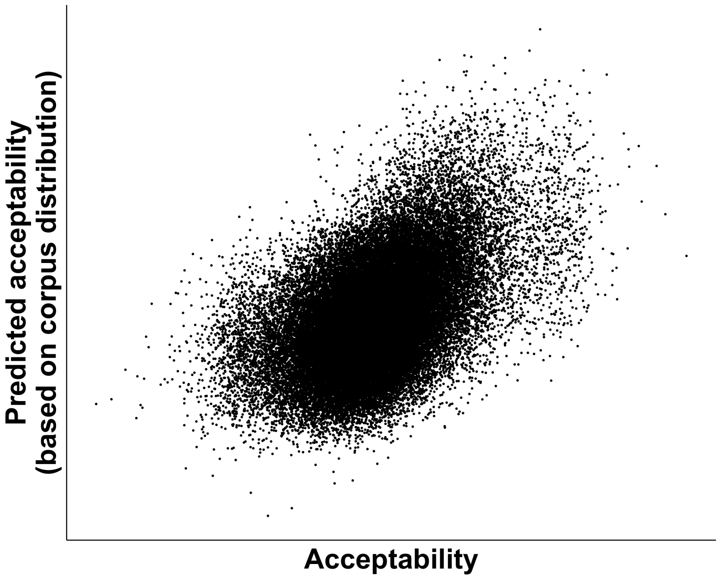
1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements
5. Whether there is a clausal complement, and if so...
   5.1 ...what the complementizer is (if any)
   5.2 ...what the WH word is (if any)
   5.3 ...what the subject is (if any)

Features extracted see White 2015 for details

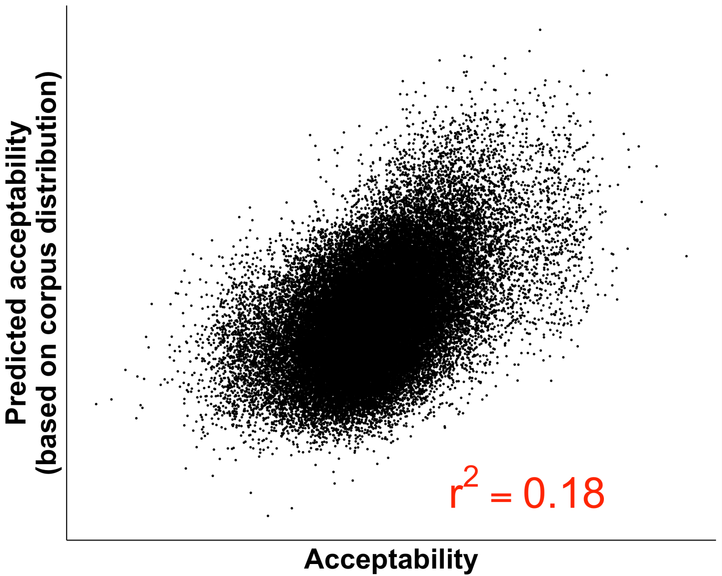
1. Form of the matrix subject (i.e. potentially expletive?)
2. Tense/aspect for matrix verb (and all matrix auxiliaries)
3. Whether there is direct or indirect NP objects
4. Whether there are other PP complements
5. Whether there is a clausal complement, and if so...
    5.1 ...what the complementizer is (if any)
    5.2 ...what the WH word is (if any)
    5.3 ...what the subject is (if any)
    5.4 ...tense/aspect for the embedded verb (and all auxiliaries)

Acceptability v. PukWaC corpus counts

Acceptability v. PukWaC corpus counts



Acceptability v. PukWaC corpus counts



Question

Is this r^2 good enough?

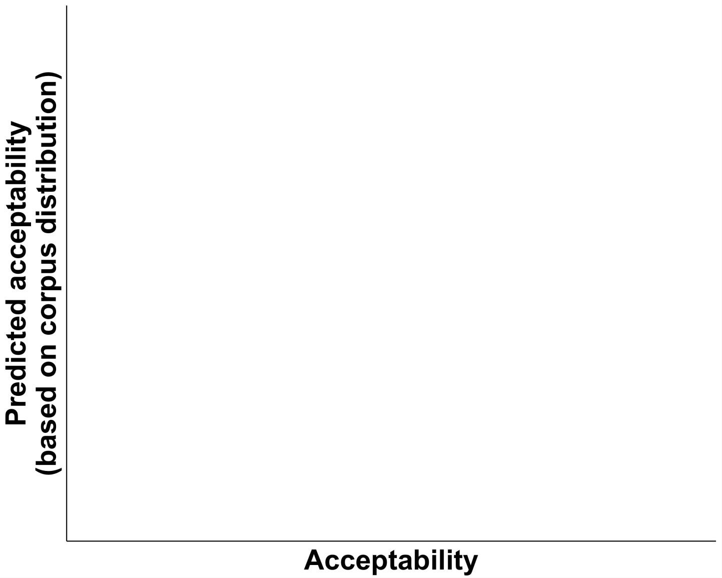
Question

Is this r^2 good enough?

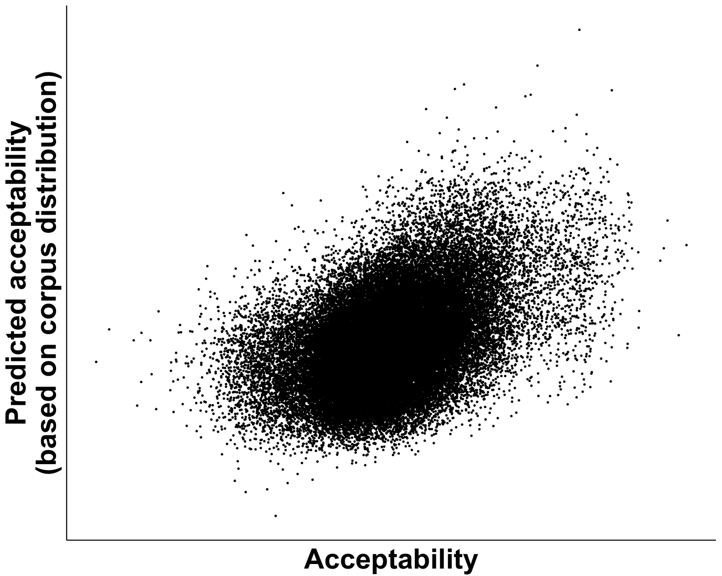
Non-answer

Better than existing alternatives, such as VALEX (Korhonen et al. 2006)

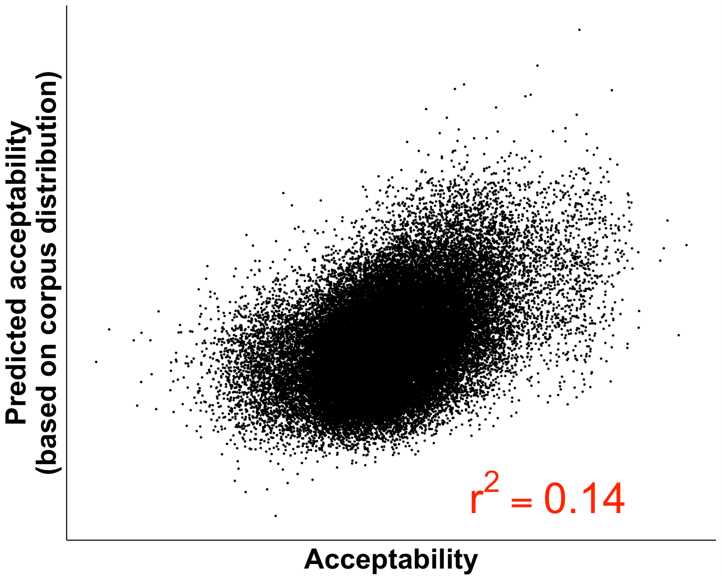
Acceptability v. VALEX corpus counts



Acceptability v. VALEX corpus counts



Acceptability v. VALEX corpus counts



Predicting acceptability

Question

Is this r^2 good enough?

Non-answer

Better than existing alternatives, such as VALEX (Korhonen et al. 2006)

Predicting acceptability

Question

Is this r^2 good enough?

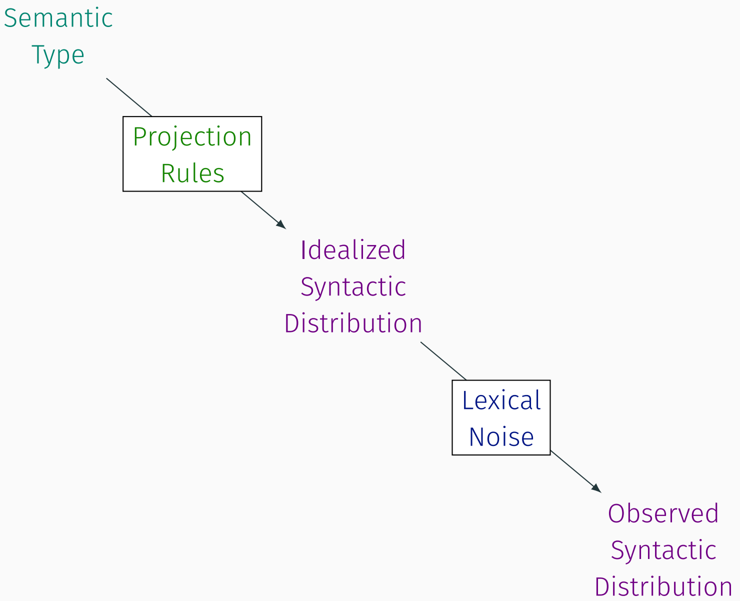
Non-answer

Better than existing alternatives, such as VALEX (Korhonen et al. 2006)

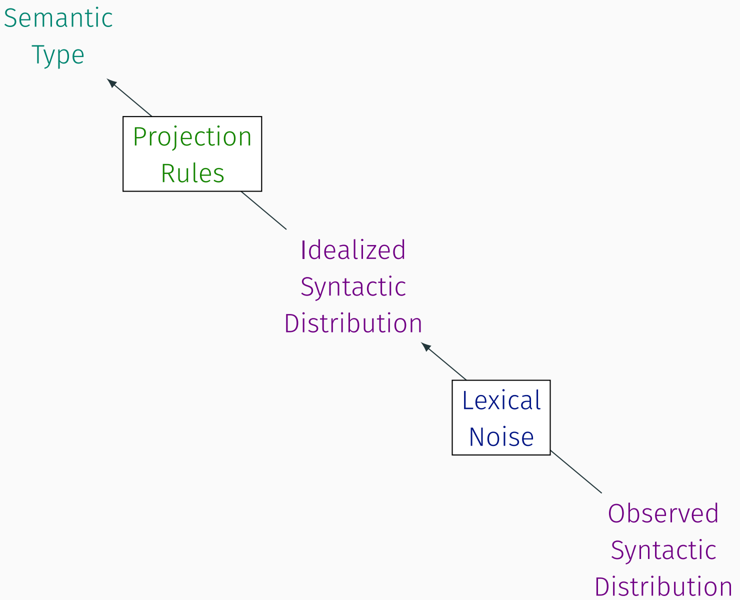
Possible answer

Maybe if the noise model is set up correctly.

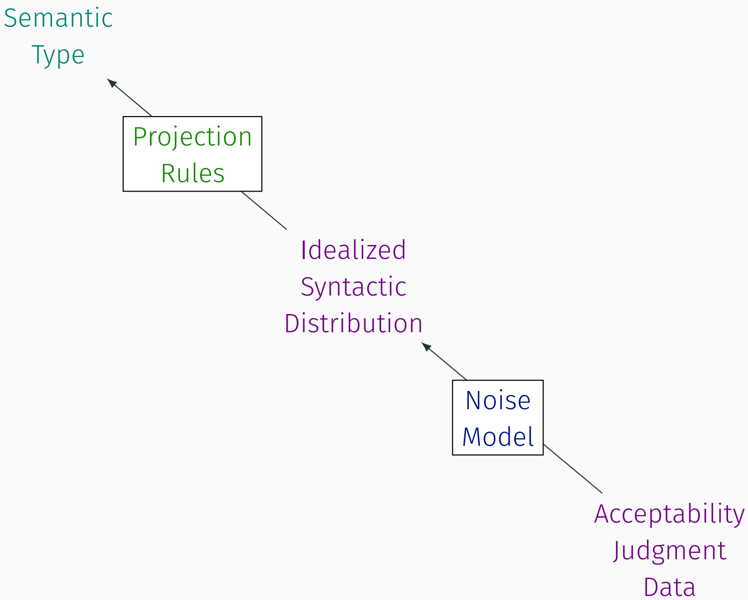
A model of S-selection and projection



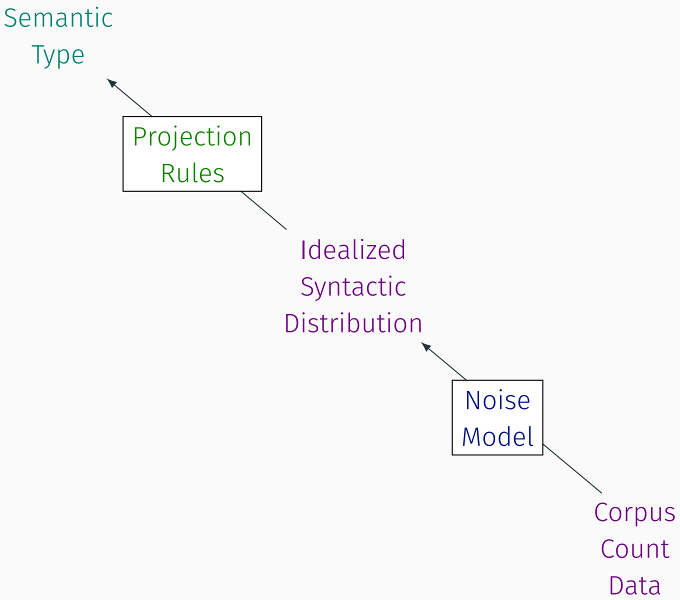
A model of S-selection and projection



A model of S-selection and projection



A model of S-selection and projection



Core model

Keep model of S-selection and projection constant.

Fitting the model

Core model

Keep model of S-selection and projection constant.

Noise model

Negative binomial mixed effects model (Church & Gale 1995, Gelman et al. 2013)

Fitting the model

Core model

Keep model of S-selection and projection constant.

Noise model

Negative binomial mixed effects model (Church & Gale 1995, Gelman et al. 2013)

Algorithm

Adam optimizer (Kingma & Ba 2014)

Fitting the model

Selecting a number of type signatures

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

Selecting a number of type signatures

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

Result

24 is the optimal number of type signatures according to AIC

Fitting the model

Selecting a number of type signatures

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

Result

24 is the optimal number of type signatures according to AIC

Reporting findings

Compare count model with 24 type signatures to acceptability model with 12

Predicting acceptability

Question

Is this r^2 good enough?

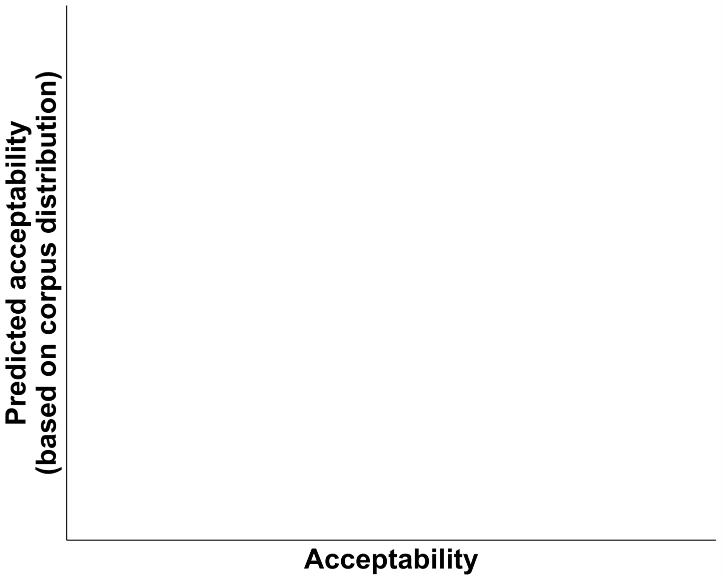
Non-answer

Better than existing alternatives, such as VALEX (Korhonen et al. 2006)

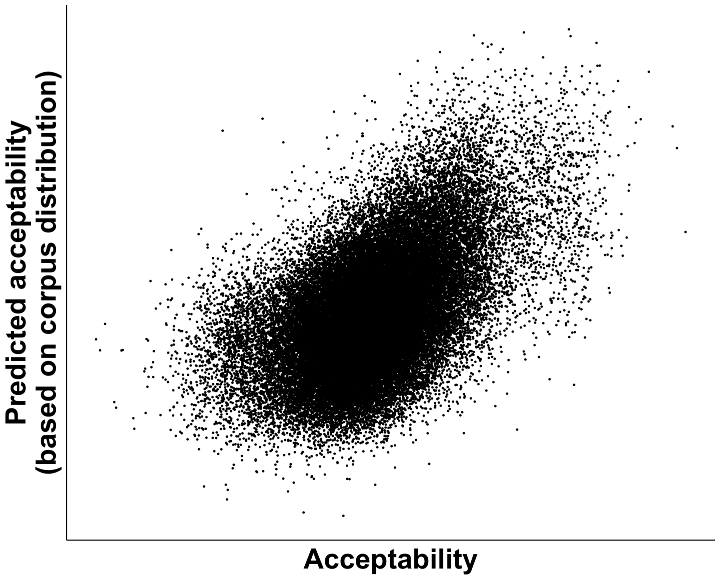
Possible answer

Maybe if the noise model is set up correctly.

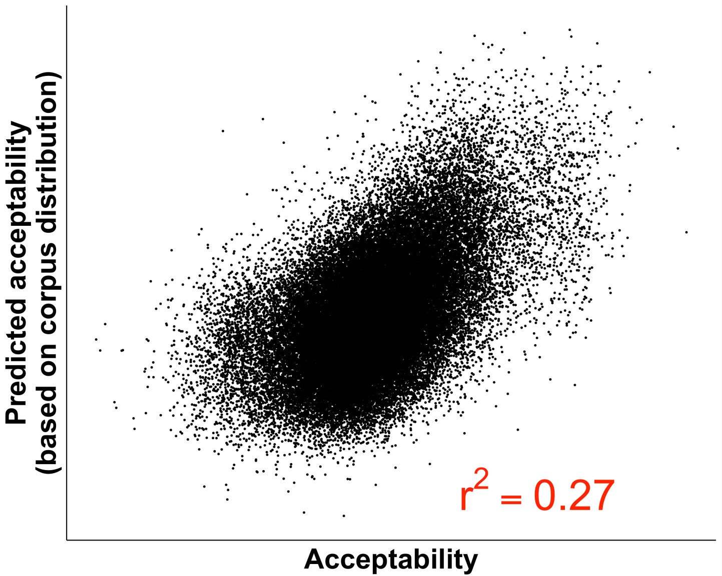
Acceptability v. VALEX corpus counts



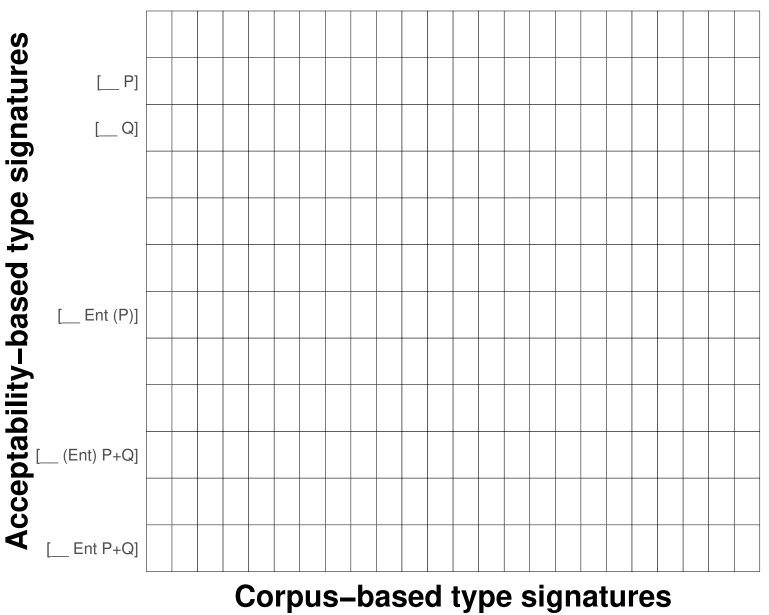
Acceptability v. VALEX corpus counts



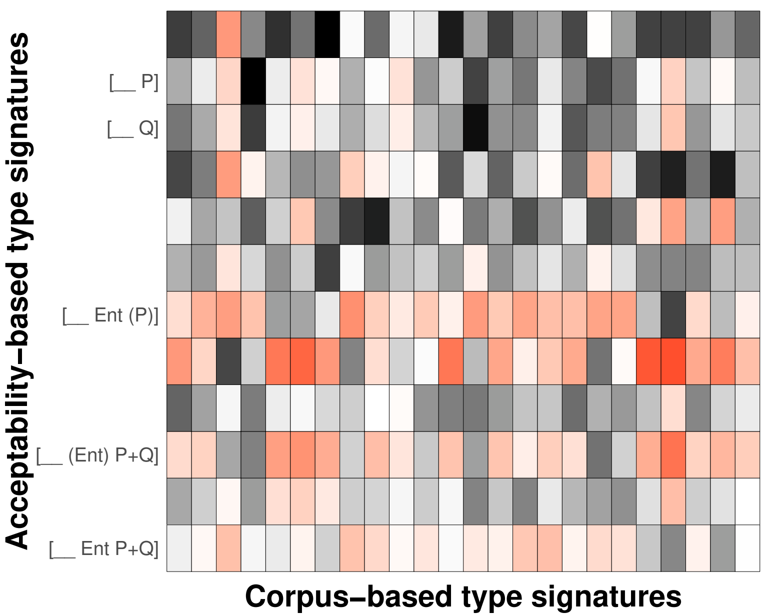
Acceptability v. VALEX corpus counts



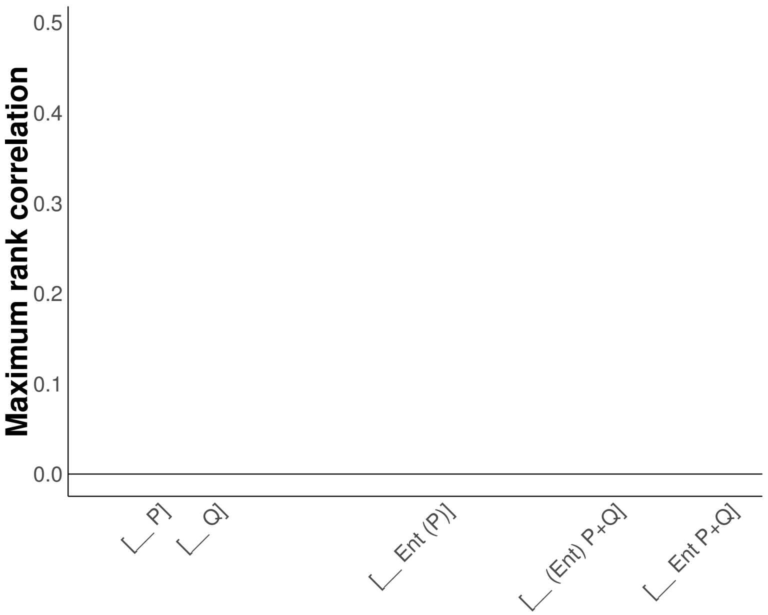
Acceptability v. corpus type signatures



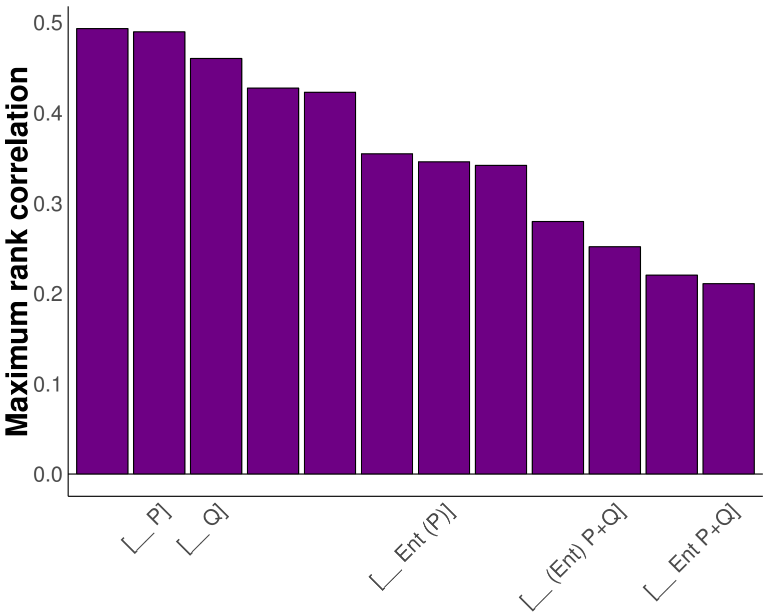
Acceptability v. corpus type signatures



Acceptability v. corpus type signatures



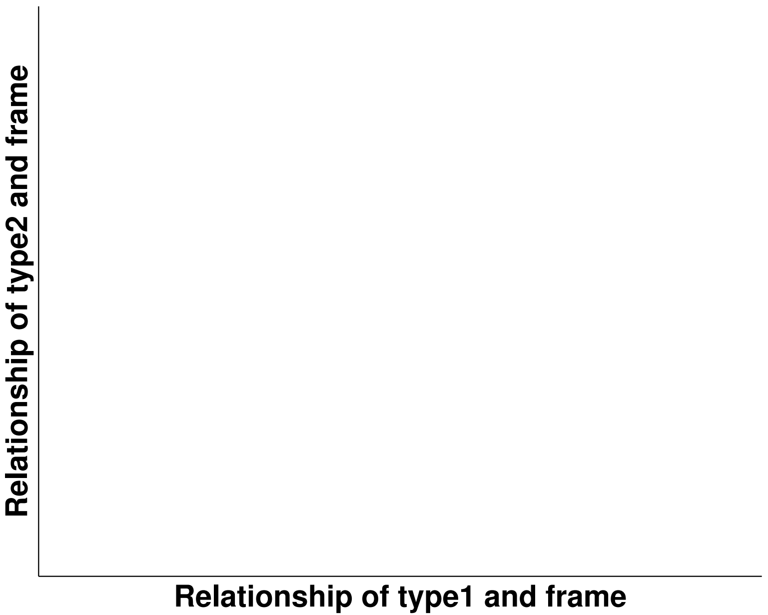
Acceptability v. corpus type signatures



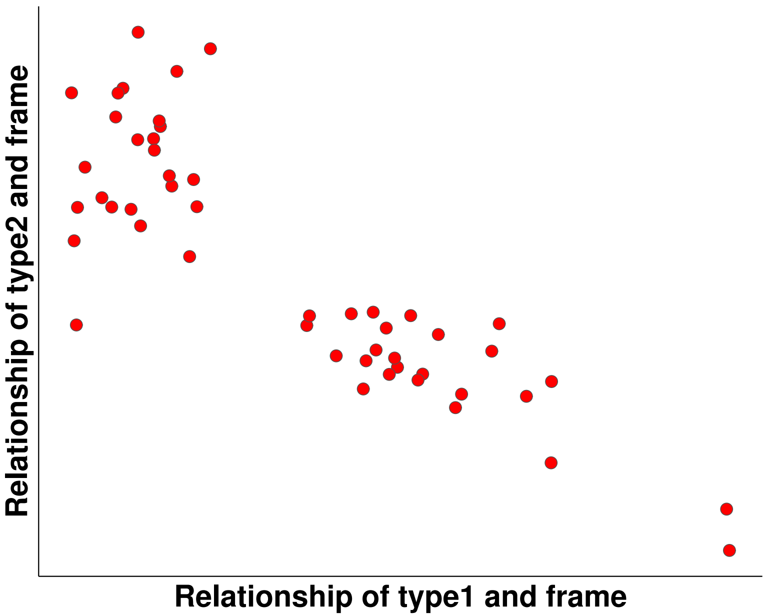
Question

What do the closest corpus type signatures to [___Ent P⊕Q] and [___(Ent) P⊕Q] look like?

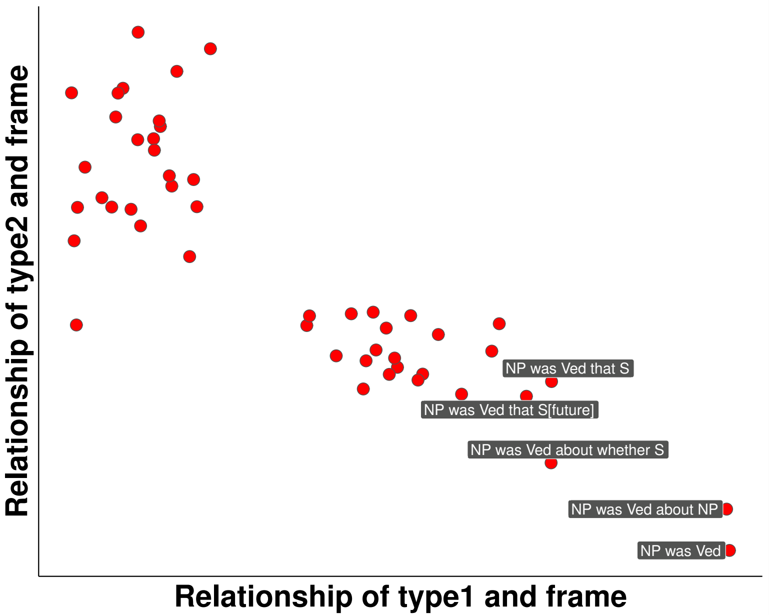
Recipients in the corpus type signatures



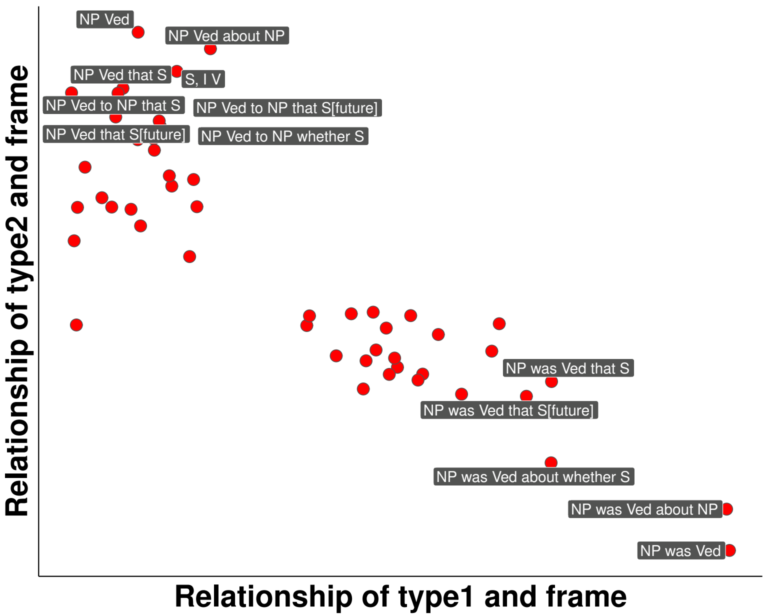
Recipients in the corpus type signatures



Recipients in the corpus type signatures



Recipients in the corpus type signatures



Question

What do the closest corpus type signatures to [___Ent P⊕Q] and [___(Ent) P⊕Q] look like?

Acceptability v. corpus type signatures

Question

What do the closest corpus type signatures to [___Ent P$\oplus$Q] and [___(Ent) P$\oplus$Q] look like?

Question

What do the closest corpus type signatures to [___Ent P$\oplus$Q] and [___(Ent) P$\oplus$Q] look like?

Findings

Shared type signatures

[___P] and [___Q] show up as separate type signatures in both the acceptability solution and the corpus solution

Findings

Shared type signatures

[___P] and [___Q] show up as separate type signatures in both the acceptability solution and the corpus solution

Differing type signatures

[___Ent P⊕Q] and [___(Ent) P⊕Q] only show up in the acceptability solution

Question #1

Why would the communicative type signatures not be found in the corpus?

Interim discussion

Question #1

Why would the communicative type signatures not be found in the corpus?

Potential answer

The corpus data is enough to tell that the predicate is communicative, but you need to know that communicatives take P⊕Q

Question #2

What about the other 18 type signatures?

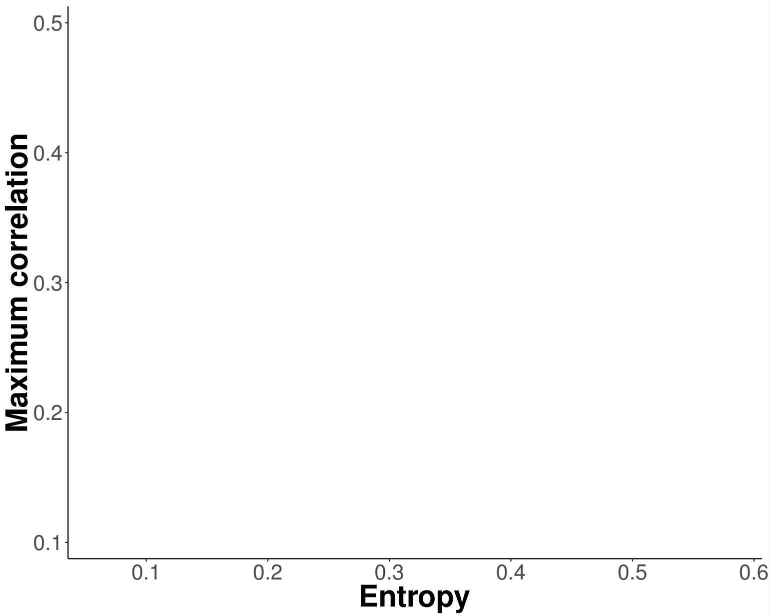
Question #2

What about the other 18 type signatures?

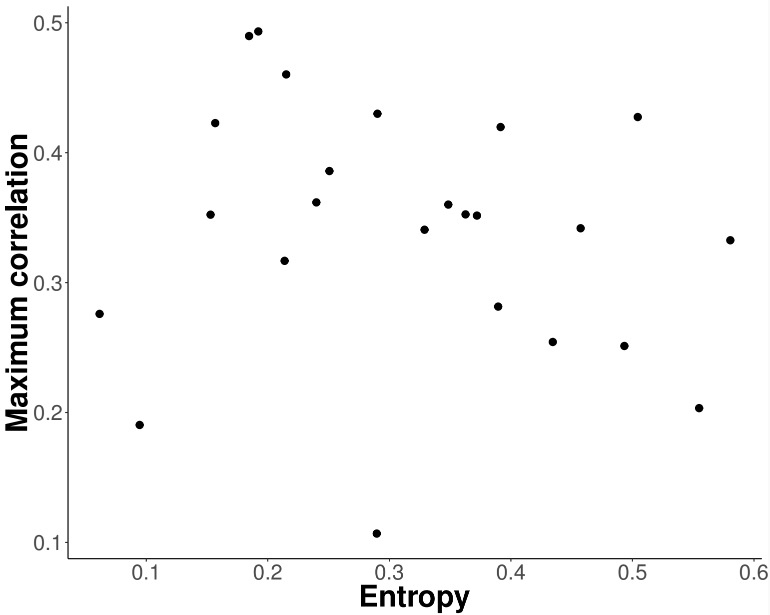
Potential answer

These tend to be junk, but we may be able to filter them out by looking at how uncertain the model is that particular verbs take that type signature overall (measured using entropy).

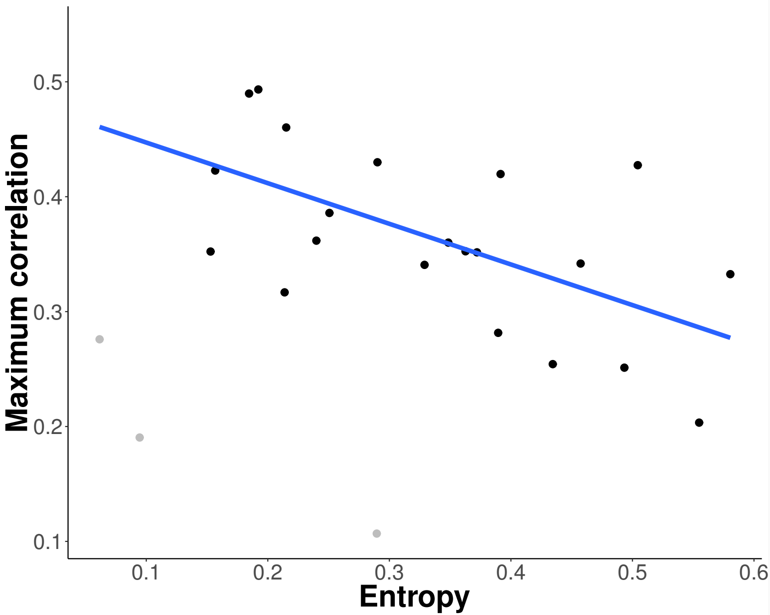
Interim discussion



Interim discussion



Interim discussion



Conclusions and future directions

Conclusions

Structure of the domain

What types of things do predicates relate?

Conclusions

Structure of the domain

What types of things do predicates relate?

S(emantic)-selection

Which predicates relate which types of things?

Conclusions

Structure of the domain

What types of things do predicates relate?

S(emantic)-selection

Which predicates relate which types of things?

Projection rules

What is the mapping from those types to syntactic structures?

Conclusion

Main contribution

A **computational method** for **scaling distributional analysis** that is agnostic about the form of the distribution.

Conclusion

Case study

Responsive predicates: take both interrogative and declaratives

(7) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

Case study

Responsive predicates: take both interrogative and declaratives

(7) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

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

Case study

Responsive predicates: take both interrogative and declaratives

(7) a. John knows {that, whether} it's raining.
b. John told Mary {that, whether} it was raining.

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

Finding #1

Cognitives take separate P and Q types, while communicatives take a hybrid P$\oplus$Q type.

Conclusion

Case study

Responsive predicates: take both interrogative and declaratives

(7) a. John knows {that, whether} it's raining.
    b. John told Mary {that, whether} it was raining.

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

Finding #1

Cognitives take separate P and Q types, while communicatives take a hybrid P⊕Q type.

Finding #2

Only the cognitive types are replicated when looking at a corpus (though apparent communicative types still show up).

Further investigation of type signatures

Seven other type signatures that are also remarkably coherent

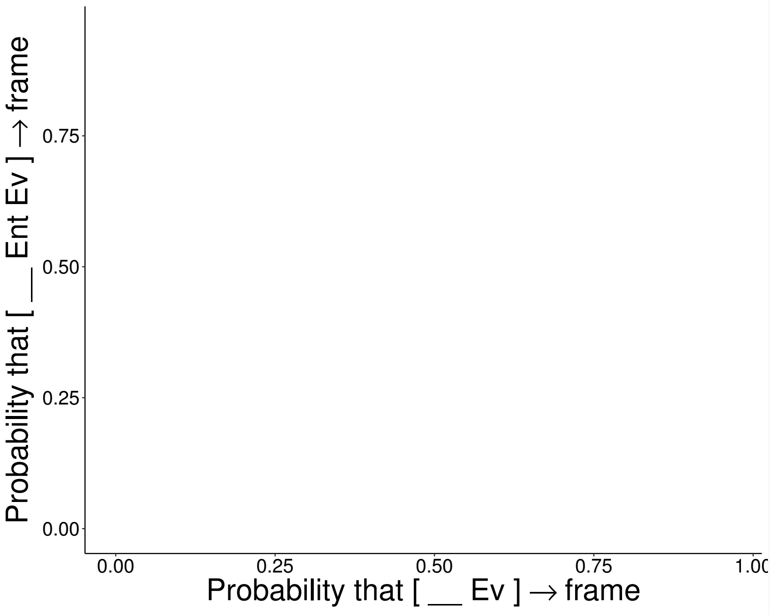
Further investigation of type signatures

Seven other type signatures that are also remarkably coherent

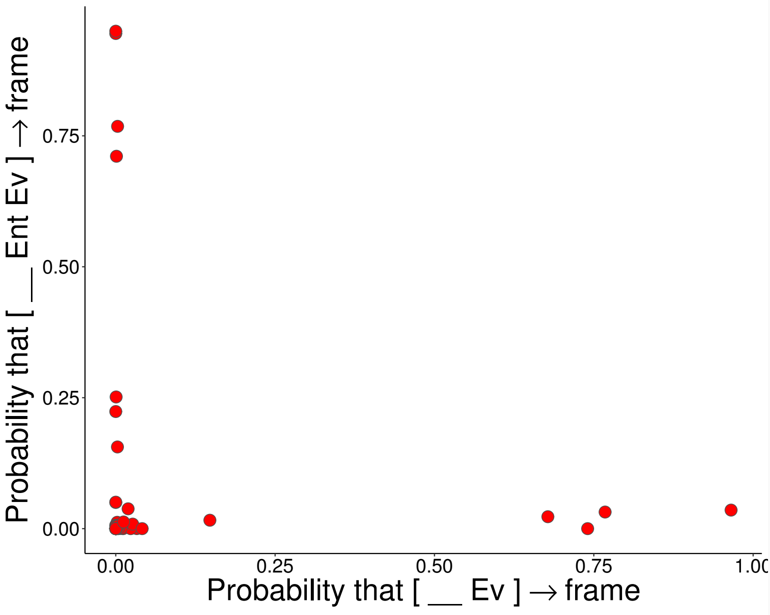
Example

Many nonfinite-taking verbs

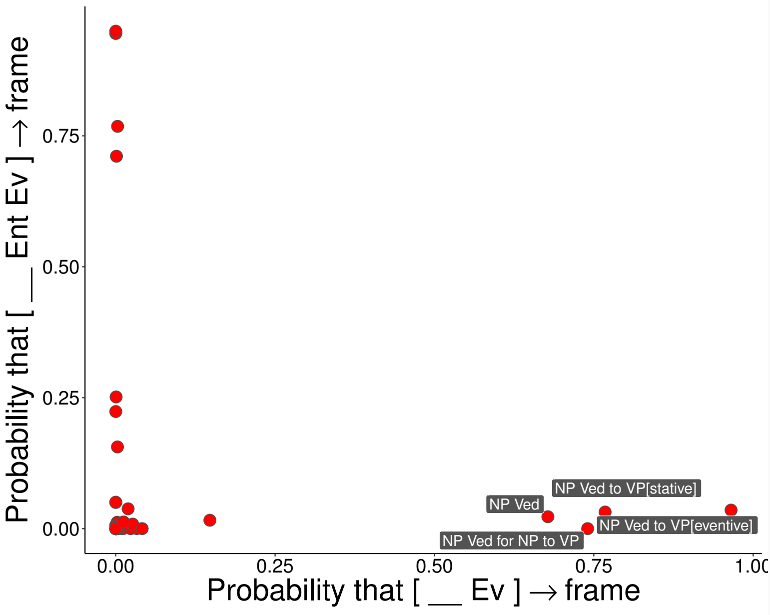
Projection: events



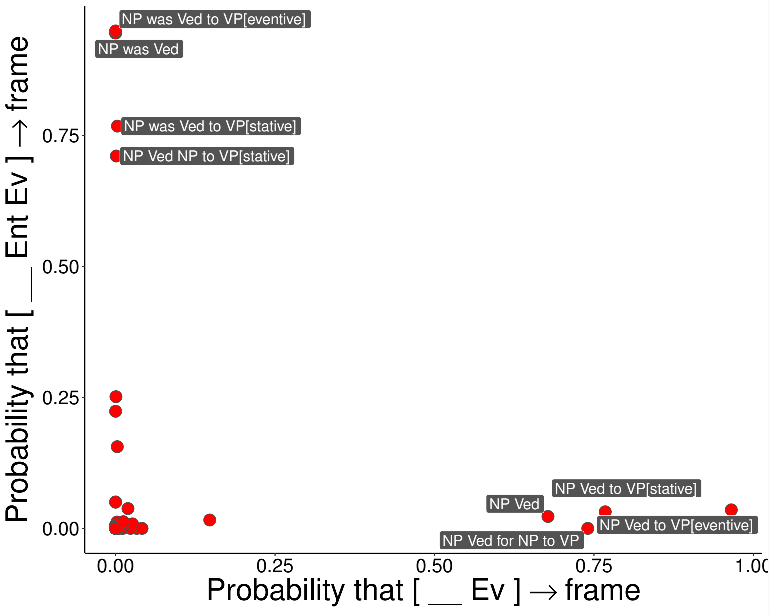
Projection: events



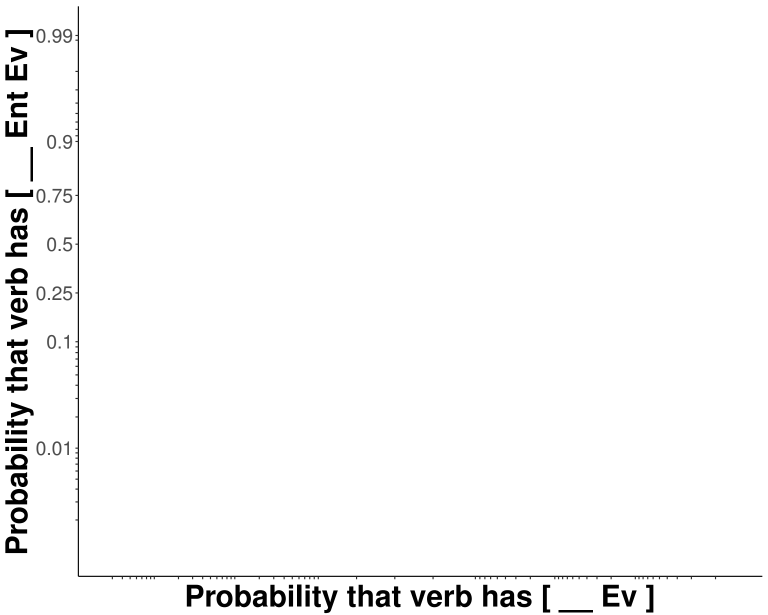
Projection: events



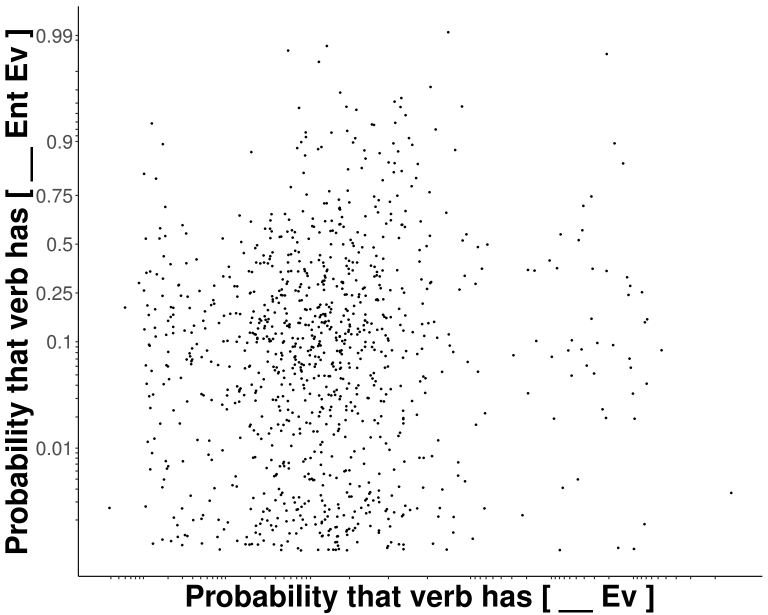
Projection: events



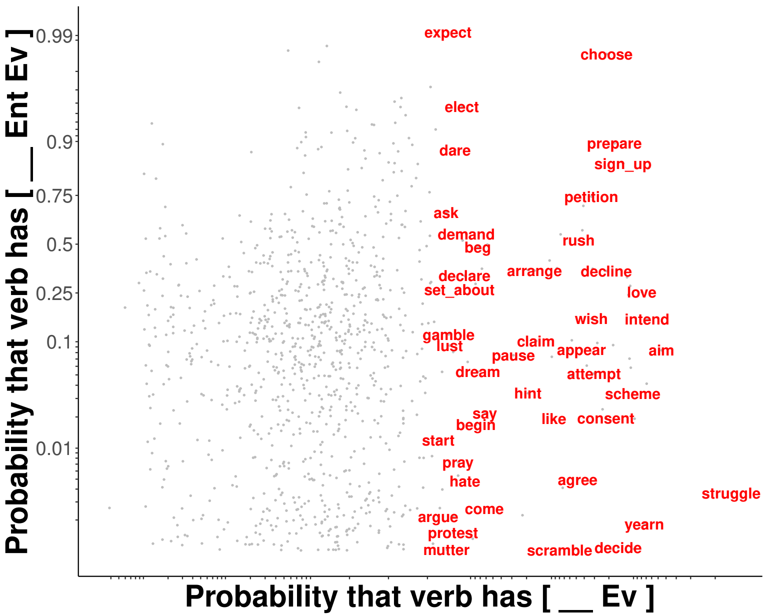
S-selection: events



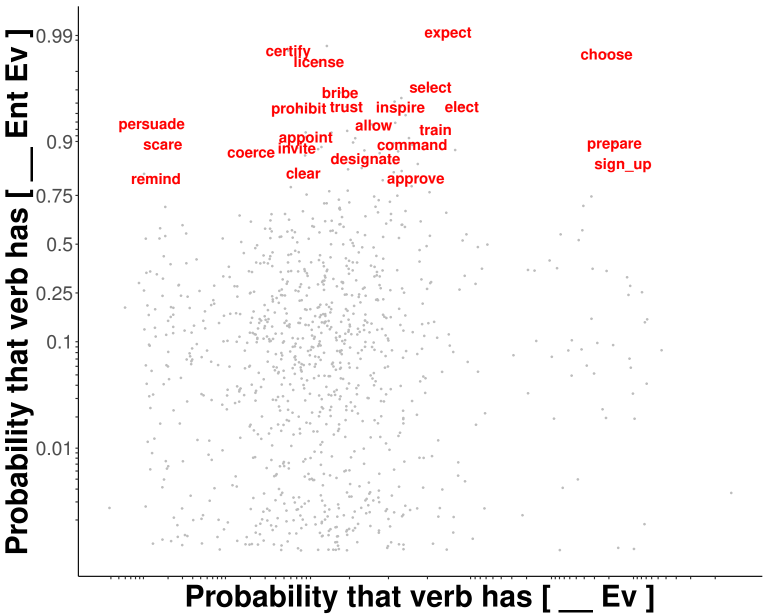
S-selection: events



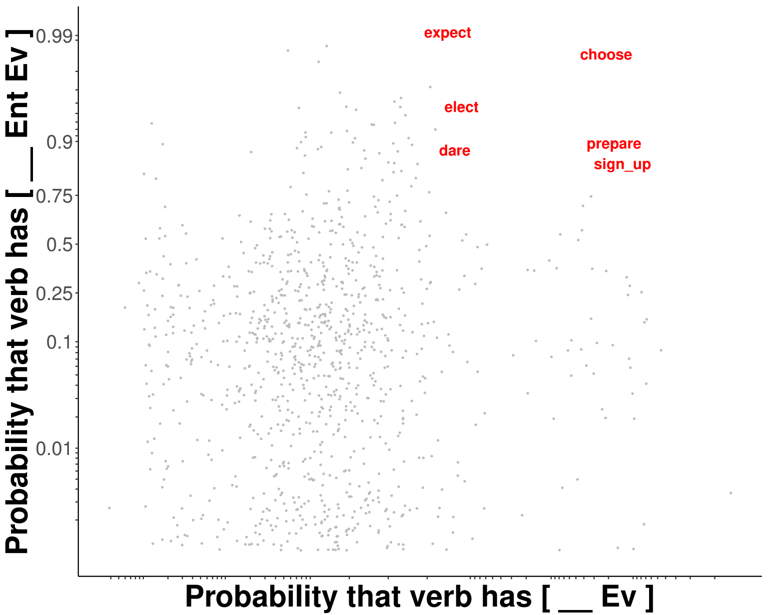
S-selection: events



S-selection: events



S-selection: events



Future directions

Atomic v. structured type signatures

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

Idea

Build a model that represents mappings from...

1. ...verbs to the primitive types they relate
2. ...type signatures to the primitive types they are constituted of
3. ...primitive types to the syntactic constituents they map to

Homophony v. regular polysemy v. underspecification

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

Future directions

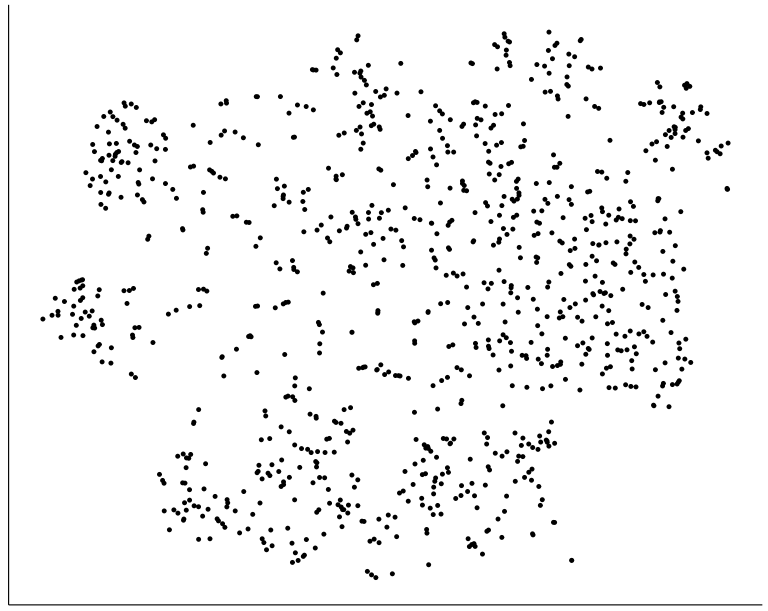
Homophony v. regular polysemy v. underspecification

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

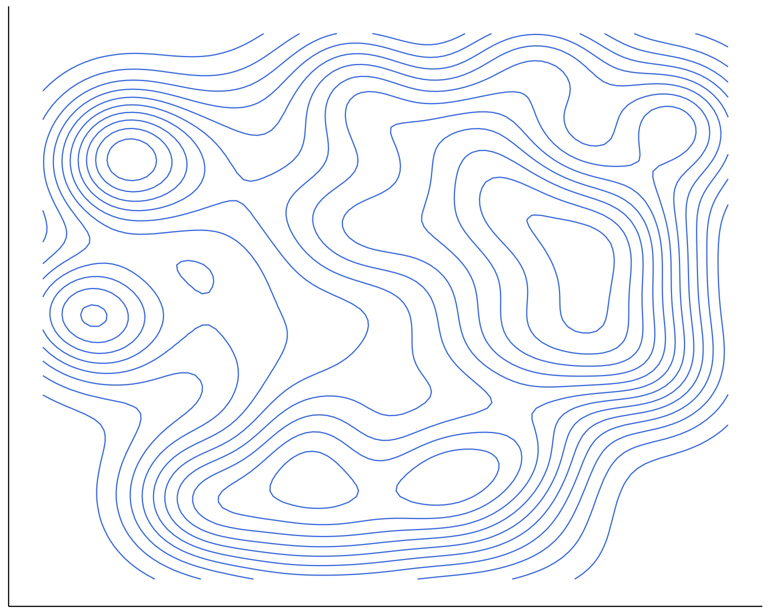
Idea

Polysemous verbs are those that fall outside dense regions of type signature space.

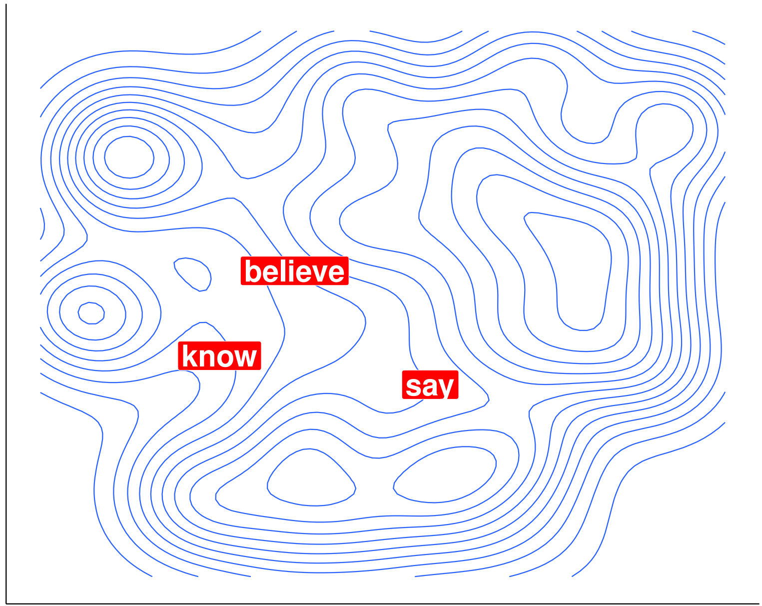
Finding polysemous verbs



Finding polysemous verbs



Finding polysemous verbs



Future directions

Homophony v. regular polysemy v. underspecification

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

Idea

Polysemous verbs are those that fall outside dense regions of type signature space.

Future directions

Homophony v. regular polysemy v. underspecification

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

Idea

Polysemous verbs are those that fall outside dense regions of type signature space.

Question

Can we learn rules of regular polysemy using an elaborated version of the model proposed here?

Thanks

I am grateful to audiences at Johns Hopkins University, SALT 26, and ESSLLI 2017 for discussion of this work. I would like to thank Ben Van Durme, Shevaun Lewis, and Dee Reisinger in particular for useful comments.

This work was funded in part by 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.

Thanks

Some of the broader ideas also developed with...

Valentine Hacquard

University of Maryland

Department of Linguistics

Jeff Lidz

University of Maryland

Department of Linguistics

Agresti, Alan. 2014. *Categorical Data Analysis*. John Wiley & Sons.

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

Aloni, Maria & Floris Roelofsen. 2011. Interpreting concealed questions. *Linguistics and Philosophy* 34(5). 443–478.

Baker, Carl Leroy. 1968. *Indirect Questions in English*: University of Illinois dissertation.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation* 43(3). 209–226.

Carter, Richard. 1976. Some linking regularities. In *On Linking: Papers by Richard Carter* Lexicon Project Working Papers (Vol. 25), Cambridge, MA: MIT Center for Cognitive Science.

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

Church, Kenneth W. & William A. Gale. 1995. Poisson mixtures. *Natural Language Engineering* 1(02). 163–190.

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. 2010a. *Concealed Questions. In Search of Answers*: University of Massachusetts, Amherst dissertation.

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

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari & Donald B. Rubin. 2013. *Bayesian data analysis*. CRC press.

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.

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

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

Bibliography IV

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. *Studies on the Semantics of Questions and the Pragmatics of Answers*: University of Amsterdam dissertation.

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. 1979. Concealed questions. In R. Bäuerle, U. Egli & A.v. Stechow (eds.), *Semantics from Different Points of View* Springer Series in Language and Communication, 51–60. Springer.

Bibliography V

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for know. In *Proceedings of Israel Association for Theoretical Linguistics*, 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.

Kingma, Diederik & Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Korhonen, Anna, Yuval Krymolowski & Ted Briscoe. 2006. A large subcategorization lexicon for natural language processing applications. In *Proceedings of LREC*, vol. 6, .

Lahiri, Utpal. 2002. *Questions and Answers in Embedded Contexts*. Oxford University Press.

Levin, Beth. 1993. *English Verb Classes and Alternations: A preliminary investigation*. Chicago: University of Chicago Press.

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

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.

Nivre, Joakim, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov & Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(02). 95–135.

Pesetsky, David. 1982. *Paths and Categories*: Massachusetts Institute of Technology 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 Todd Snider (ed.), *Semantics and Linguistic Theory*, vol. 23, 336–357.

Romero, Maribel. 2005. Concealed questions and specificational subjects. *Linguistics and Philosophy* 28(6). 687–737.

Spector, Benjamin & Paul Egré. 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*: Massachusetts Institute of Technology 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.

White, Aaron Steven & Kyle Rawlins. 2016. A computational model of S-selection. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Semantics and Linguistic Theory*, vol. 26, 641–663.

White, Aaron Steven & Kyle Rawlins. 2017. Question agnosticism and change of state. In *Proceedings of Sinn und Bedeutung 21*, to appear.