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
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(semantic)-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**.
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...
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.
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
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
Focus

Syntactic distribution of ~1000 English clause-embedding verbs
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?
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?
Idea (≈ 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 (≈ 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 (≈ 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.

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.


Finding #1 (based on acceptability judgments)
Different answer for communicative and cognitive verbs.
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.


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

Introduction

A model of S-selection & projection

Acceptability dataset
  Data collection
  Model fitting and results

Corpus Dataset
  Data collection
  Model fitting and results

Conclusions and future directions
A model of S-selection & projection
Many verbs are **syntactically multiplicitous**

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

---

14
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]$_{NP}$.

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

What do the projection rules look like? How are a verb's semantic type signatures projected onto its syntactic type signatures (subcategorization frames)?

Grimshaw's notation

\[
\langle \langle \langle s,t \rangle, t \rangle, t \rangle = \begin{cases} Q \end{cases}
\]

Montagovian notation

\[
S = \begin{cases} NP \end{cases}
\]

Semantic type

Syntactic type
Projection

What do the projection rules look like? How are a verb's semantic type signatures projected onto its syntactic type signatures (subcategorization frames)?


\[
\begin{align*}
\text{Semantic type} & : [\_\_Q] \\
\text{Projection} & : [\_\_S] \\
\{Projection}{Projecting} & \Rightarrow \text{Syntactic type} : [\_\_NP]
\end{align*}
\]

(Grimshaw's notation)

Projection

Semantic type

Projection

Syntactic type

[___S]

⟨⟨⟨s,t⟩,t⟩, t⟩

(Montagovian notation)

[___NP]
What do the projection rules look like?


Semantic type

Projection

Syntactic type

[___S]  [___Q]  [___NP]
A model of S-selection and projection

Semantic Type

Projection Rules

Syntactic Distribution

Lexical Noise

Model Construction

Lexicon Constructor

Projection Constructor

Acceptability Judgment

Data Corpus

Corpus Count

Projection Rules
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)

\[
S\text{-selection} \circ \text{projection} \lor C\text{-selection} = \text{syntactic distribution}
\]
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**)  

\[ S\text{-selection} \circ \text{projection} \lor \text{C-selection} = \text{syntactic distribution} \]

The multiplicative approach (Pesetsky 1982, 1991)
Verbs are related to **semantic type signatures** (**S-selection**); **C-selection** is an epiphenomenon of verbs’ **abstract case**  

\[ S\text{-selection} \circ \text{projection} \land \text{case} = \text{syntactic distribution} \]
Two kinds of lexical idiosyncrasy

Shared core see White & Rawlins 2016 for formal details

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

$S$-selection $\circ$ projection $\otimes$ noise = syntactic distribution
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

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A boolean model of S-selection

know → \{[\_P], [\_Q]\}
A boolean model of S-selection

\[ \text{know} \rightarrow \{[\_P], [\_Q]\} \quad \text{wonder} \rightarrow \{[\_Q]\} \]
A boolean model of S-selection

think $\rightarrow \{[\_\_P]\}$  
know $\rightarrow \{[\_\_P], [\_\_Q]\}$  
wonder $\rightarrow \{[\_\_Q]\}$
A boolean model of S-selection

think → \{[\_\_P]\}    know → \{[\_\_P], [\_\_Q]\}    wonder → \{[\_\_Q]\}

\[
S = \begin{pmatrix}
    \_\_P & \_\_Q & \ldots \\
    \text{think} & 1 & 0 & \ldots \\
    \text{know} & 1 & 1 & \ldots \\
    \text{wonder} & 0 & 1 & \ldots \\
    \ldots & \vdots & \vdots & \ldots \\
\end{pmatrix}
\]
A boolean model of projection

\[ \text{\[\text{P}\mapsto\{\text{that S}, \text{NP}, \ldots}\]} \quad \text{\[\text{Q}\mapsto\{\text{whether S}, \text{NP}, \ldots}\]} \]
A boolean model of projection

\[
\begin{align*}
[\_\_P] & \rightarrow \{[\_\_\text{that } S], [\_\_\text{NP}], \ldots\} & [\_\_Q] & \rightarrow \{[\_\_\text{whether } S], [\_\_\text{NP}], \ldots\}
\end{align*}
\]

\[
\Pi = \begin{bmatrix}
[\_\_P] & 1 & 0 & 1 & \cdots \\
[\_\_Q] & 0 & 1 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]
A boolean model of idealized syntactic distribution

\[
\hat{D}(\text{VERB, SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} S(\text{VERB, } t) \land \Pi(t, \text{SYNTYPE})
\]
A boolean model of idealized syntactic distribution

\[ \hat{D}(\text{VERB, SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} S(\text{VERB, } t) \land \Pi(t, \text{SYNTYPE}) \]

\[
\begin{pmatrix}
\text{think} & \begin{pmatrix} 1 & 0 & \ldots \end{pmatrix} & \begin{pmatrix} 1 \end{pmatrix} & 0 & 1 & \ldots \\
\text{know} & \begin{pmatrix} 1 & 1 & \ldots \end{pmatrix} & \begin{pmatrix} 0 \end{pmatrix} & 1 & 1 & \ldots \\
\text{wonder} & \begin{pmatrix} 0 & 1 & \ldots \end{pmatrix} & \ldots & \ldots & \ldots & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\]
A boolean model of idealized syntactic distribution

\[
\hat{D}(\text{VERB, SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} S(\text{VERB}, t) \wedge \Pi(t, \text{SYNTYPE})
\]
A boolean model of idealized syntactic distribution

\[ \hat{D}(\text{know, \_\_that S}) = \bigvee_{t \in \{\_\_P, \_\_Q, \ldots\}} S(\text{know, t}) \land \Pi(t, \_\_that S) \]

<table>
<thead>
<tr>
<th>__P</th>
<th>__Q</th>
<th>__that S</th>
<th>__whether S</th>
<th>__NP</th>
<th>__NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{think, \_\_that S}) = \bigvee_{t \in \{\_\_P, \_\_Q, \ldots\}} S(\text{think, t}) \land \Pi(t, \_\_that S) \]

<table>
<thead>
<tr>
<th>__P</th>
<th>__Q</th>
<th>__that S</th>
<th>__whether S</th>
<th>__NP</th>
<th>__NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{know, \_\_that S}) = \bigvee_{t \in \{\_\_P, \_\_Q, \ldots\}} S(\text{know, t}) \land \Pi(t, \_\_that S) \]

<table>
<thead>
<tr>
<th>__P</th>
<th>__Q</th>
<th>__that S</th>
<th>__whether S</th>
<th>__NP</th>
<th>__NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>__NP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{think, \_\_that S}) = \bigvee_{t \in \{\_\_P, \_\_Q, \ldots\}} S(\text{think, t}) \land \Pi(t, \_\_that S) \]
A boolean model of idealized syntactic distribution

\[ \hat{D}(\text{know}, [\_\_\text{that S}]) = \bigvee_{t \in \{[\_\_\_P],[\_\_\_Q],...\}} S(\text{know}, t) \land \Pi(t, [\_\_\_\text{that S}]) \]

<table>
<thead>
<tr>
<th></th>
<th>[___P]</th>
<th>[___Q]</th>
<th>...</th>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>know</td>
<td>[___P]</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>wonder</td>
<td>[___Q]</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{wonder}, [\_\_\_NP]) = \bigvee_{t \in \{[\_\_\_P],[\_\_\_Q],...\}} S(\text{wonder}, t) \land \Pi(t, [\_\_\_\_NP]) \]

<table>
<thead>
<tr>
<th></th>
<th>[___P]</th>
<th>[___Q]</th>
<th>...</th>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>wonder</td>
<td>[___P]</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td></td>
<td>[___Q]</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{think}, [\_\_\_\_P]) = \bigvee_{t \in \{[\_\_\_P],[\_\_\_Q],...\}} S(\text{think}, t) \land \Pi(t, [\_\_\_\_\_P]) \]

<table>
<thead>
<tr>
<th></th>
<th>[___P]</th>
<th>[___Q]</th>
<th>...</th>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td></td>
<td>[___P]</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td></td>
<td>[___Q]</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{D}(\text{know}, [\_\_\_\_P]) = \bigvee_{t \in \{[\_\_\_P],[\_\_\_Q],...\}} S(\text{know}, t) \land \Pi(t, [\_\_\_\_\_P]) \]

<table>
<thead>
<tr>
<th></th>
<th>[___P]</th>
<th>[___Q]</th>
<th>...</th>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td></td>
<td>[___P]</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td></td>
<td>[___Q]</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A boolean model of idealized syntactic distribution

\[ \hat{D}(\text{wonder, [___NP]})) = \bigvee_{t \in \{[___P],[___Q],\ldots\}} S(\text{wonder, } t) \land \Pi(t, [___NP]) \]

<table>
<thead>
<tr>
<th></th>
<th>[___P]</th>
<th>[___Q]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

...
A model of S-selection and projection

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A boolean model of observed syntactic distribution

\[ \forall t \in \text{SYNTYPE} : \mathbf{D}(\text{wonder}, t) = \hat{\mathbf{D}}(\text{wonder}, t) \otimes \mathbf{N}(\text{wonder}, t) \]
A boolean model of observed syntactic distribution

\[ \forall t \in \text{SYNTYPE} : D(\text{wonder}, t) = \hat{D}(\text{wonder}, t) \otimes N(\text{wonder}, t) \]

\[
\begin{pmatrix}
\text{think} & 1 & 0 & 1 & \cdots \\
\text{know} & 1 & 1 & 1 & \cdots \\
\text{wonder} & 0 & 1 & 1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix} = 
\begin{pmatrix}
\text{think} & 0 & 0 & 0 & \cdots \\
\text{know} & 0 & 0 & 0 & \cdots \\
\text{wonder} & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]
A boolean model of observed syntactic distribution

∀t ∈ SYNTYPE : D(wonder, t) = \hat{D}(wonder, t) \otimes N(wonder, t)
A boolean model of observed syntactic distribution

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

<table>
<thead>
<tr>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>[___that S]</th>
<th>[___whether S]</th>
<th>[___NP]</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>know</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A model of S-selection and projection

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution

Corpus Count Data

Acceptability Judgment
Acceptability dataset
Data available at megaattitude.com
Ordinal (1-7 scale) acceptability ratings
Ordinal (1-7 scale) acceptability ratings for 1000 clause-embedding verbs
Verb selection
Ordinal (1-7 scale) acceptability ratings

for

1000 clause-embedding verbs

×

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

Syntactic type

- NP
- PP
- S

Frame construction

Syntactic type

NP → PP

PP → S

S
Frame construction

Syntactic type

- NP
- PP
- S

ACTIVE
PASSIVE

that
for
whether
which
would
to
-ed
-ing

+Q
-FIN
+FIN

36
Frame construction

Syntactic type

- NP
  - ACTIVE
  - PASSIVE
- PP
  - COMP
  - TENSE
- S
Frame construction

- Syntactic type
  - NP
  - PP
  - S

- ACTIVE
- PASSIVE
- COMP
- TENSE

- that [+Q]
- for
- ∅
Frame construction

Syntactic type

NP
ACTIVE
PASSIVE

PP

S
COMP

TENSE

that [+Q]
for ∅

whether
which NP
Frame construction

Syntactic type

NP
ACTIVE  PASSIVE

PP

S

COMP

TENSE

[-FIN]  [+FIN]

[∅]  [for]  [+Q]

which NP  whether

[∅]  that
Frame construction

Syntactic type

NP
ACTIVE PASSIVE

PP

S

COMP

TENSE

[+FIN] [-FIN]

that [+]Q for ∅ [+FIN] [-FIN]

whether which NP -ed would
Frame construction

Syntactic type

NP

ACTIVE

PASSIVE

PP

S

COMP

TENSE

that [+Q]

for ∅ [+FIN]

whether [−FIN]

which NP

−ed

would

to ∅

−ing
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. \( \text{know} + \text{NP V} \{\text{that, whether}\} \text{ S} \)
   Someone knew \{that, whether\} something happened.
b. \( \text{tell} + \text{NP V NP} \{\text{that, whether}\} \text{ S} \)
   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

Examples of responsives

(6)

a. \( \text{know} + \text{NP V \{that, whether\} S} \)
   Someone knew \{that, whether\} something happened.

b. \( \text{tell} + \text{NP V NP \{that, whether\} S} \)
   Someone told someone \{that, whether\} something happened.
Data collection

- 1,000 verbs × 50 syntactic frames = 50,000 sentences
Data collection

• 1,000 verbs × 50 syntactic frames = 50,000 sentences
• 1,000 lists of 50 items each
Data collection

- 1,000 verbs × 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 × 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 × 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

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

The diagram shows a scatter plot with the x-axis labeled as "NP__ed S" and the y-axis labeled as "NP__ed whether S". The data points are plotted across a range from 1 to 7 on both axes.
Results
Results

NP __ed S
NP __ed whether S

know
want

1 2 3 4 5 6 7
Results

<table>
<thead>
<tr>
<th>NP ed whether S</th>
<th>NP __ed S</th>
</tr>
</thead>
<tbody>
<tr>
<td>wonder</td>
<td></td>
</tr>
<tr>
<td>know</td>
<td></td>
</tr>
<tr>
<td>think</td>
<td></td>
</tr>
<tr>
<td>want</td>
<td></td>
</tr>
</tbody>
</table>
A model of S-selection and projection

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A model of S-selection and projection

- Semantic Type
- Projection Rules
  - Idealized Syntactic Distribution
    - Lexical Noise
      - Observed Syntactic Distribution
A model of S-selection and projection

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Noise Model

Acceptability Judgment Data
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
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

\[
\hat{D}(\text{VERB, SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} S(\text{VERB}, t) \land \Pi(t, \text{SYNTYPE})
\]

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>...</th>
<th>that S</th>
<th>whether S</th>
<th>NP</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>think</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>know</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>wonder</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\[
\hat{D}(\text{VERB, SYNTYPE}) = \bigvee_{t \in \text{SEMTYPES}} S(\text{VERB}, t) \land \Pi(t, \text{SYNTYPE})
\]
A boolean model of idealized syntactic distribution

\[
\hat{D}(\text{know}, \text{[___ that S]}) = 1 - \prod_{t \in \{\begin{array}{c}
[\_P] \\
[\_Q]
\end{array}\}} 1 - S(\text{know}, t) \times \Pi(t, \text{[___ that S]})
\]
Wrapping with probabilities

\[
\mathbb{P}(S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}]) = \mathbb{P}(S_{\text{VERB}, t})\mathbb{P}(\Pi[t, \text{SYNTYPE}] \mid S_{\text{VERB}, t}) \\
= \mathbb{P}(S_{\text{VERB}, t})\mathbb{P}(\Pi[t, \text{SYNTYPE}])
\]

\[
\mathbb{P} \left( \bigvee_t S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}] \right) = \mathbb{P} \left( \neg \bigwedge_t \neg(S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}]) \right) \\
= 1 - \mathbb{P} \left( \bigwedge_t \neg(S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}]) \right) \\
= 1 - \prod_t \mathbb{P}(\neg(S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}])) \\
= 1 - \prod_t 1 - \mathbb{P}(S_{\text{VERB}, t} \land \Pi[t, \text{SYNTYPE}]) \\
= 1 - \prod_t 1 - \mathbb{P}(S_{\text{VERB}, t})\mathbb{P}(\Pi[t, \text{SYNTYPE}])
\]
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 \( \text{(Agresti 2014)} \)

Algorithm
Adam optimizer (basically, fancy gradient descent) \( \text{(Kingma & Ba 2014)} \)

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)
High-level idea

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

\[
\begin{array}{c}
[\_\_P] \\
\downarrow \\
[\_\_\text{that S}] \\
\downarrow \\
[\_\_\text{whether S}]
\end{array}
\quad
\begin{array}{c}
[\_\_Q] \\
\downarrow \\
[\_\_\text{that S}] \\
\downarrow \\
[\_\_\text{whether S}]
\end{array}
\]
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
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

[___P]  [___Q]
   \downarrow    \downarrow
[___that S]  [___whether S]
   \downarrow    \downarrow
[___(Ent) P\oplus Q]
   \downarrow    \downarrow
[___to NP that S]  [___to NP whether S]
Findings

\[ \begin{array}{cc}
\text{\[ \_\_\, P \_\_ ]} & \text{\[ \_\_\, Q \_\_ ]} \\
\downarrow & \downarrow \\
\text{\[ \_\_\, \text{that } S \] } & \text{\[ \_\_\, \text{whether } S \] }
\end{array} \]

\[ \_\_\, (\text{Ent})\, P \oplus Q \_\_ \]

\[ \begin{array}{cc}
\text{\[ \_\_\, \text{to NP that } S \] } & \text{\[ \_\_\, \text{to NP whether } S \] } \\
\end{array} \]
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: propositions and questions

Probability that [__ P] → frame

Probability that [__ Q] → frame
Projection: propositions and questions
Projection: propositions and questions

NP Ved NP
NP Ved S
S, I V
NP Ved that S
NP Ved that S [future]
NP Ved whether S
NP Ved whether S [future]
NP Ved whether to VP
NP Ved whichNP S
NP Ved whichNP to VP

Probability that [ __ P ] → frame
Probability that [ __ Q ] → frame

NP Ved that S
NP Ved that S [future]
NP Ved NP
NP Ved whichNP S
NP Ved whichNP to VP
NP Ved whether S
NP Ved whether S [future]
S-selection: propositions and questions

Probability that verb has [ __ P ]

Probability that verb has [ __ Q ]
S-selection: propositions and questions
S-selection: propositions and questions

Probability that verb has [ __ P ]

Probability that verb has [ __ Q ]

0.001
0.01
0.1
0.25
0.5
0.75
0.9
0.99

0.001

Probability that verb has [ __ P ]

Probability that verb has [ __ Q ]

acknowledge
admit
agree
allege
announce
assume
attest
believe
bet
conclude
decide
deduce
determine
discover
expect
explain
fear
find_out
gather
guarantee
hear
hope
know
point_out
predict
sense
speculate
swear
take
think
warrant
wish

S-selection: propositions and questions

Probability that verb has [ __ Q ]

Probability that verb has [ __ P ]
S-selection: propositions and questions

Probability that verb has [ __ P ]

Probability that verb has [ __ Q ]

- assume
- decide
- deduce
- explain
- find_out
- know
- point_out
- sense
- deduce
- explain
- find_out
- know
- assume
- point_out
Findings

\[
\text{[\_\_P]} \quad \text{[\_\_Q]}
\]

\[
\downarrow \quad \downarrow
\]

\[
\text{[\_\_that S]} \quad \text{[\_\_whether S]}
\]
Projection: optional recipients

Probability that \[ \underline{\text{P} + \text{Q}} \] → frame

Probability that \[ \underline{\text{P}} \] → frame
Projection: optional recipients
Projection: optional recipients

NP Ved S
NP Ved that S
NP Ved that S[future]

Probability that \[ \___ P \] \(\rightarrow\) frame

Probability that \[ \___ (Rec) P+Q \] \(\rightarrow\) frame
Projection: optional recipients

```
NP Ved NP
NP Ved S
S, I V
NP Ved that S
NP Ved that S [future]
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]
```

Probability that \([\_ \_ P + Q ] \rightarrow \text{frame}\)

Probability that \([\_ \_ (\text{Rec}) P + Q ] \rightarrow \text{frame}\)
S-selection: optional recipients

Probability that verb has [__ (Ent) P+Q]

Probability that verb has [__ P]
S-selection: optional recipients
S-selection: optional recipients

Probability that verb has [ __ P ]

Probability that verb has [ __ (Ent) P+Q ]

acknowledge
admit
agree
announce
assume
attest
believe
claim
conclude
confirm
decide
deduce
deny
detect
determine
doubt
dream
expect
explain
fear
feel
gather
hint
hope
infer
know
maintain
point_out
recall
reveal
restate
reveal
say
speculate
trust
understand
verify

Probability values:
0.0001
0.001
0.01
0.1
0.25
0.5
0.75
0.9
0.99
S-selection: optional recipients

- Probability that verb has \[ \text{__ (Ent) P+Q} \]
- Probability that verb has \[ \text{__ P} \]

- Verb examples:
  - lie
  - share
  - fess_up
  - type
  - holler
  - radio
  - report
  - infer
  - recall
  - hint
  - acknowledge
  - reveal
  - confirm
  - deny
  - point_out
  - admit
  - boast
  - bullshit
  - broadcast
  - divulge
  - mention
  - clarify
  - explain

Probabilities:
- 0.0001
- 0.001
- 0.01
- 0.1
- 0.25
- 0.5
- 0.75
- 0.9
- 0.99
S-selection: optional recipients

- Probability that verb has \[ \_ \ (\text{Ent}) P+Q \]
- Probability that verb has \[ \_ P \]

verbs:
- acknowledge
- admit
- announce
- confirm
- deny
- explain
- hint
- infer
- point_out
- recall
- reveal
- verify

probabilities:
- 0.0001
- 0.001
- 0.01
- 0.1
- 0.25
- 0.5
- 0.75
- 0.9
- 0.99

Coordinate system:
- Probability that verb has \[ \_ \ (\text{Ent}) P+Q \]
- Probability that verb has \[ \_ P \]
Findings

\[
\begin{align*}
[\_\_ P] & \quad [\_\_ Q] \\
\downarrow & \quad \downarrow \\
[\_\_ \text{that } S] & \quad [\_\_ \text{whether } S] \\
& \quad (\text{Ent}) \quad P \oplus Q \\
[\_\_ \text{to NP that } S] & \quad [\_\_ \text{to NP whether } S]
\end{align*}
\]
Projection: obligatory recipients/experiencers
Projection: obligatory recipients/experiencers

Probability that [ __ Ent (P) ] $\rightarrow$ frame

Probability that [ __ Ent P+Q ] $\rightarrow$ frame
Projection: obligatory recipients/experiencers

- NP was Ved about NP
- NP was Ved about whether S
- NP was Ved that S
- NP was Ved that S\[future\]

The graph shows the probability that certain constructions are used as a function of the probability that a specific frame is associated with the sentence. The x-axis represents the probability that the sentence has frame (P), and the y-axis represents the probability that a certain construction is used. The data points indicate a positive correlation between the two variables, with higher probabilities of the construction being used corresponding to higher probabilities of the frame being associated with the sentence.
Projection: obligatory recipients/experiencers

Probability that [ ____ Ent (P) ] \rightarrow frame

[red points on a graph]

NP Ved NP that S[future]
NP Ved NP that S
NP Ved NP whether S[future]
NP Ved NP whether S
NP was Ved whichNP to VP
NP was Ved whether to VP
NP was Ved whether S[future]
NP was Ved that S[future]
NP was Ved that S
S-selection: obligatory recipients/experiencers

Probability that verb has [ __ Ent P+Q ]

Probability that verb has [ __ Ent (P) ]
S-selection: obligatory recipients/experiencers

Probability that verb has [ __ Ent P+Q ]

Probability that verb has [ __ Ent (P) ]
S-selection: obligatory recipients/experiencers

Probability that verb has [ __ Ent P+Q ]

Probability that verb has [ __ Ent (P) ]
S-selection: obligatory recipients/experiencers

Probability that verb has [ __ Ent P+Q ]

Probability that verb has [ __ Ent (P) ]
S-selection: obligatory recipients/experiencers

Probability that verb has $[\_\_\text{Ent (P)}]$ vs Probability that verb has $[\_\_\text{Ent P+Q}]$
Findings

\[
\begin{align*}
\text{[\_\_P]} & \quad \text{[\_\_Q]} \\
\quad & \quad \\
\text{[\_\_that S]} & \quad \text{[\_\_whether S]} \\
\quad & \quad \\
\text{[\_\_Ent P} \oplus \text{Q}] & \\
\quad & \quad \\
\text{[\_\_NP that S]} & \quad \text{[\_\_NP whether S]}
\end{align*}
\]
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 $P \oplus Q$ is reducible to lexical semantics?
Interim discussion

Acceptability of $CP[+Q]$ vs. Veridicality
Interim discussion

Acceptability of \[ CP[+Q] \] vs. Veridicality

Graph showing a positive correlation between the acceptability of \[ CP[+Q] \] and its veridicality.
Interim discussion

Veridicality
Acceptability of $[\text{CP}[+Q]]$

![Scatter plot showing the relationship between acceptability and veridicality.](96)
Question

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

Is there anything to say about whether selection for $P$, $Q$, or $P \oplus Q$ 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 $P \oplus Q$ 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)
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)

I was amazed that they had come.

ROOT

SUBJ

VC

OBJ

SUBJ

VC

SENT

NN

VBD

VVN

IN

PPS

VHD

VVN

.
Subcategorization frame extraction

Features extracted see White 2015 for details

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

Predicted acceptability (based on corpus distribution)

Acceptability

$r^2 = 0.18$
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

Predicted acceptability (based on corpus distribution)

Acceptability
Acceptability v. VALEX corpus counts

Predicted acceptability (based on corpus distribution) vs. Acceptability
Acceptability v. VALEX corpus counts

$\text{Predicted acceptability (based on corpus distribution)}$

$\text{Acceptability}$

$r^2 = 0.14$
Question

Is this $r^2$ good enough?

Non-answer

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

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A model of S-selection and projection

Semantic Type

Projection Rules

Idealized Syntactic Distribution

Lexical Noise

Observed Syntactic Distribution
A model of S-selection and projection
A model of S-selection and projection

Semantic
Type

Projection
Rules

Idealized
Syntactic
Distribution

Noise
Model

Corpus
Count
Data
Fitting the model

Core model

Keep model of \textit{S-selection} and \textit{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)
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
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

Predicted acceptability (based on corpus distribution) vs. Acceptability
Acceptability v. VALEX corpus counts

\[ r^2 = 0.27 \]
Acceptability v. corpus type signatures

Acceptability-based type signatures

Corpus-based type signatures
Acceptability v. corpus type signatures
Acceptability v. corpus type signatures

Maximum rank correlation
Acceptability v. corpus type signatures

Maximum rank correlation

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>Q</th>
<th>Ent (P)</th>
<th>(Ent) P+Q</th>
<th>Ent P+Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Question

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

| Relationship of type2 and frame | Relationship of type1 and frame |
Recipients in the corpus type signatures

Relationship of type1 and frame
Relationship of type2 and frame
Recipients in the corpus type signatures

Relationship of type1 and frame

- NP was Ved about NP
- NP was Ved about whether S
- NP was Ved that S
- NP was Ved that S[future]

Relationship of type2 and frame
Recipients in the corpus type signatures

Relationship of type1 and frame

Relationship of type2 and frame

NP Ved
NP Ved about NP
NP Ved that S
S, I V
NP Ved to NP that S
NP Ved to NP that S[future]
NP Ved that S[future]
NP Ved to NP whether S
NP Ved to NP that S[future]
NP Ved whether S
NP was Ved that S
NP was Ved that S[future]
NP was Ved about whether S
NP was Ved about NP
NP was Ved
Question

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

What do the closest corpus type signatures to \( [\_\_\text{Ent} P \oplus Q] \) and \( [\_\_\text{(Ent)} P \oplus Q] \) look like?

Question

What do the closest corpus type signatures to \( [\_\_\text{Ent} P \oplus Q] \) and \( [\_\_\text{(Ent)} P \oplus Q] \) 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.
Interim discussion

Question #1

Why would the communicative type signatures not be found in the corpus?
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 $P \oplus Q$
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

![Graph](image)

- Maximum correlation
- Entropy

The graph illustrates the relationship between maximum correlation and entropy.
Interim discussion

![Graph showing the relationship between entropy and maximum correlation. The graph includes a line of best fit.](image)
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(semantic)-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?
Main contribution

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

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.


Finding #1

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


Finding #1

Cognitives take separate $P$ and $Q$ types, while communicatives take a hybrid $P \oplus Q$ 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
Future directions

Further investigation of type signatures
Seven other type signatures that are also remarkably coherent

Example
Many nonfinite-taking verbs
Projection: events

![Graph showing the probability of events given frames]

- Probability that [ __ Ent Ev ] → frame
- Probability that [ __ Ev ] → frame
Projection: events

![Graph showing probability distributions for different event types.](image)
Projection: events

Probability that [ __ Ent Ev ] → frame

Probability that [ __ Ev ] → frame
S-selection: events

Probability that verb has [ __ Ev ]

Probability that verb has [ __ Ent Ev ]
S-selection: events

Probability that verb has [__ Ent Ev]

Probability that verb has [__ Ev]
S-selection: events

Probability that verb has [ __ Ev ]

Probability that verb has [ __ Ent Ev ]

agree
aimappear
argue
arrange
ask
attempt
beg
begin
choose
claim
come
consent
dare
decide
decide
demand
dream
delect
dare
prepare
sign_up
petition
rush
dare
declare
set_about
expect
choose
elect

0.01
0.1
0.25
0.5
0.75
0.9
0.99
S-selection: events

Probability that verb has [ __ Ev ]

Probability that verb has [ __ Ent Ev ]
S-selection: events

Probability that verb has [ __ Ev ]

Probability that verb has [ __ Ent Ev ]

0.01
0.1
0.25
0.5
0.75
0.9
0.99

choose
dare
elect
expect
prepare
sign_up
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
Future directions

Homophony v. regular polysemy v. underspecification

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

believe
know
say
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.
Some of the broader ideas also developed with...

Valentine Hacquard  
*University of Maryland*  
*Department of Linguistics*

Jeff Lidz  
*University of Maryland*  
*Department of Linguistics*


White, Aaron Steven. 2015. Information and Incrementality in Syntactic Bootstrapping: University of Maryland dissertation.

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