A computational model of S-selection [*]

Aaron Steven White
Johns Hopkins University

Kyle Rawlins
Johns Hopkins University

Abstract We develop a probabilistic model of S(emantic)-selection that encodes both the notion of systematic mappings from semantic type signature to syntactic distribution—i.e., projection rules—and the notion of selectional noise—e.g., C(ategory)-selection, L(exical)-selection, and/or other independent syntactic processes. We train this model on data from a large-scale judgment study assessing the acceptability of 1,000 English clause-taking verbs in 50 distinct syntactic frames, finding that this model infers coherent semantic type signatures. We focus in on type signatures relevant to interrogative and declarative selection, arguing that our results suggest a principled split between cognitive verbs, which select distinct proposition and question types, and communicative verbs, which select a single hybrid type.

Keywords: semantic selection, attitude verbs, experimental syntax, computational semantics

1 Introduction

It has long been assumed that a verb's syntactic distribution is determined by at least two kinds of lexical information: (i) the verb's semantic type signatures and (ii) its morphosyntactic features (Chomsky 1965; Gruber 1965; Fillmore 1970; Zwicky 1971; Jackendoff 1972; Carter 1976; Pinker 1984, 1989; Grimshaw 1990; Levin 1993; Chomsky 1973; Bresnan 1972; Grimshaw 1979; Pesetsky 1982, 1991 among others). The first of these is often termed S(emantic)-selection (Grimshaw 1979; Pesetsky 1982, 1991); the second goes under various names, though perhaps the most neutral term is subcategorization (Chomsky 1965). Standard distributional analyses in the theoretical literature have had tremendous success in uncovering the nature of S-selection and its relationship to the syntax—i.e., projection rules. But as theories scale to the entire lexicon, these approaches hit a limit, imposed by the sheer size of lexica and by bounds on human analysts' memory and processing power.

In an effort to address these limitations, we develop a combined experimental-computational methodology for studying the selectional behavior of verbs at the

---

scale of the entire lexicon. The driving idea behind this method is to algorithmically implement the procedures by which human analysts form generalizations regarding selection and then apply these algorithms to large-scale data sets. To implement this idea, we first formalize and abstract existing theoretical methods for investigating selection, and then we map this abstraction into a computationally implemented inference algorithm that can operate on data of arbitrary size. We demonstrate our model's use by applying it to a notoriously challenging problem in lexical semantics: understanding S-selection in the domain of clause-embedding verbs.

To detail how an analyst might use our model's results, we focus in on a particularly tough problem in this already challenging domain: explaining the distribution of declarative and interrogative complements. A major question in this domain is why superficially quite similar verbs, like *believe*, *wonder*, and *know*, show such different distributional behavior with respect to these types of complements.

(1)   a.   John believed {that, *whether} Mary went to the store.
        b.   John wondered {*that, whether} Mary went to the store.
        c.   John knew {that, whether} Mary went to the store.

This question has a long history in the semantics literature (Karttunen 1977a,b; Groenendijk & Stokhof 1984; Heim 1994; Ginzburg 1995; Ginzburg & Sag 2001; Lahiri 2002; George 2011; Uegaki 2012, 2015a,b; Rawlins 2013), with many potential answers. Our aim is to relate our modeling results to these proposals and to argue that current proposals likely need to be relativized to particular verb classes, split in particular between cognitive and communicative verbs.

The remainder of the paper is laid out as follows. In Section 2, we review the two classic architectural proposals regarding the relationship between S-selection and syntactic distribution laid out in Grimshaw 1979, 1981 and Pesetsky 1982, 1991. In Section 3, we report on our large-scale experiment, which we use as training data for our model. In Section 4, we describe our model, which embeds an abstraction of Grimshaw's and Pesetsky's architectures, and we fit it to the experimental data reported on in Section 3. In Section 5, we analyze a portion of the modeling results, relating these results to proposals regarding the distribution of declarative and interrogative complements. And in Section 6, we conclude with remarks on how to extend the computational model beyond this case study.

2   Selection and clausal embedding

Clause-embedding verbs played a central role in much of the early literature on S-selection (Grimshaw 1979; Pesetsky 1982) and its syntactic relatives C(ategory)-selection (Bresnan 1972; Chomsky 1973) and L(exical)-selection (Pesetsky 1991, 1995). In this section, we review this literature and formalize the logical structure of

these approaches with the goal using this formalization as the basis for modeling.

2.1 Additive and multiplicative approaches

One of the central phenomena that the literature on selection must explain is the fact that semantic type signatures are not in one-to-one correspondence with syntactic type signatures (subcategorization frames). Consider the verb *ask*. *Ask* takes both interrogative complements (2a) and noun phrase complements (2b).

(2)   a.   John asked what time it was.
        b.   John asked the time.

If projection—i.e., the mapping from semantic type signatures to syntactic type signatures—were one-to-one, we would expect distinct interpretations for (2a) and (2b). But as Baker (1968) notes, there is only one interpretation: the noun phrase in (2b) is a so-called concealed question (CQ) (Heim 1979; Romero 2005; Nathan 2006; Frana 2010; Aloni & Roelofsen 2011). This might suggest that projection maps a question-selecting type signature—[_ Q]—onto two different subcategorization frames—[_ S] and [_ NP]—as schematized in (3). This pairs naturally with a range of proposals for coercing nominal expressions into question denotations.

(3)   [_ Q] $\longrightarrow$ {[_ S], [_ NP]}

One of the central puzzles in the concealed questions literature has been that a projection rule like (3) overgenerates, since not all interrogative-selecting verbs select for nominal expressions—*wonder* being a standard example.

(4)   a.   John wondered what time it was.
        b.   *John wondered the time.

The overgeneration of (4b) is problematic for a view wherein the only way to license a particular syntactic structure is by projection from a semantic type signature. Two approaches have been taken to resolving this problem, both of which rely on idiosyncratic lexical properties. Basically, it is either a listed fact about *ask* that it takes concealed questions, or it is a listed fact about *wonder* that it does not.[1]

    More specifically, the first possibility is to back the more general mapping (3) off

---

1 A third possibility is of course that there is some systematic generalization about the mapping that covers every case. For concealed questions, the closest we are aware of to complete coverage is Nathan's (2006) proposal that it is the joint selection of questions and propositions that licenses concealed questions. This proposal still undergenerates for some predicates, including *ask*. However, our goal in discussing concealed questions here is not to find the solution, but use it as an example of the key possibilities surrounding S-selection and syntactic subcategorization.

to the more specific mapping (5), thus undergenerating concealed questions for *ask*.

(5)   [_ Q] ⟶ {[_ S]}

One then needs to say how (2b) gets the interpretation it does—e.g., by stipulating a link between [_ Q] and [_ NP] in the lexical entry of *ask* itself. This approach is essentially the one that Grimshaw (1981) takes. In this system, lexical idiosyncracy is fundamentally *additive*: one starts with a syntactic distribution defined in terms of S-selection and projection rules like (5) and then adds to those distributions based on some independent lexical information, yielding the verb's observed distribution.

The second approach, introduced by Pesetsky (1982, 1991) using the mechanism of a verb's abstract case assignment, retains (3) and then lexically filters out the overgenerated (4b). In contrast to the first approach, the second is fundamentally *multiplicative* (or *intersective*): one starts with a syntactic distribution defined in terms of overgenerative projection rules like (3) and then intersects those with some independent distributional information to yield the verb's observed distribution.

Underpinning both approaches is the need to represent lexical properties (however characterized) that are unpredictable from the perspective of projection. We term such properties (*lexical*) *noise*. This term is intended to cover whatever idiosyncratic properties are employed to 'correct' the *idealized syntactic distributions* predicted by S-selection and projection alone to the *observed syntactic distribution*.

To develop the mechanics of these two approaches more clearly, we spend the next section formalizing them. We do this in two steps: first, we give a formalization in terms of function composition and set union/intersection; then, we show how to isomorphically map this formalization into one that traffics in boolean algebras. This second formalization is what we use as a base for our computational model.

2.2   A formalization of S-selection

We begin by formalizing the additive (Grimshaw-style) and multiplicative (Pesetsky-style) systems in terms of function composition and set union/intersection. Suppose we are given the function S-SELECT in (6). This function tells us that *ask* and *wonder* have only the semantic type signature [_ Q].[2]

(6) $\text{S-SELECT} = \begin{bmatrix} \text{ask} & \longrightarrow & \{[\_\ Q]\} \\ \text{wonder} & \longrightarrow & \{[\_\ Q]\} \end{bmatrix}$

Under an additive system like Grimshaw's, we define a function PROJECTION$_1$ (7a) that, when composed with S-SELECT, undergenerates CQ sentences like (2b). A

---

2 We do not intend S-SELECT as an exhaustive representation of the semantics of *ask* and *wonder*. We use it only for the purposes of illustration. For instance, *ask* likely also S-selects [_ (Ent) (Q)].

function NOISE_1 (7b) then adds in the missing parts of the DISTRIBUTION_1 (7c).[3]

(7) a. $\text{PROJECTION}_1 = \begin{bmatrix} [\_\,\text{Q}] & \longrightarrow & \{[\_\,\text{S}]\} \\ [\_\,\text{Ent}] & \longrightarrow & \{[\_\,\text{NP}]\} \end{bmatrix}$

 b. $\text{NOISE}_1 = \begin{bmatrix} \text{ask} & \longrightarrow & \{[\_\,\text{NP}]\} \\ \text{wonder} & \longrightarrow & \{\} \end{bmatrix}$

 c. $\text{DISTRIBUTION}_1 = \text{S-SELECT} \circ \text{PROJECTION}_1 \cup \text{NOISE}_1$

Formalizing a multiplicative system like Pesetsky's requires only a minor change to (7): functions PROJECTION_2 and NOISE_2 are identical in form (though not content) to PROJECTION_1 and NOISE_1, and DISTRIBUTION_2 changes the union to an intersection.

(8) a. $\text{PROJECTION}_2 = \begin{bmatrix} [\_\,\text{Q}] & \longrightarrow & \{[\_\,\text{S}], [\_\,\text{NP}]\} \\ [\_\,\text{Ent}] & \longrightarrow & \{[\_\,\text{NP}]\} \end{bmatrix}$

 b. $\text{NOISE}_2 = \begin{bmatrix} \text{ask} & \longrightarrow & \{[\_\,\text{S}], [\_\,\text{NP}]\} \\ \text{wonder} & \longrightarrow & \{[\_\,\text{S}]\} \end{bmatrix}$

 c. $\text{DISTRIBUTION}_2 = \text{S-SELECT} \circ \text{PROJECTION}_2 \cap \text{NOISE}_2$

This close formal relationship suggests a generalization of the additive and multiplicative explanations, which we take advantage of to build our model. The basic idea is to leverage the tight relationship between fields of sets and boolean algebras known from Representation Theory (Stone 1936 *et seq*) to reconstruct the additive and multiplicative approaches as boolean semiring modules. This in turn allows us to employ standard inference algorithms that can be defined over these modules.

Translating the above formalizations into boolean algebras is straightforward. The general procedure is to define for each function $f \in \{\text{S-SELECT}, \text{PROJECTION}_1, \text{NOISE}_1, \text{PROJECTION}_2, \text{NOISE}_2\}$, a matrix of indicators $\boldsymbol{M}_f$.

(9) $\boldsymbol{M}_f[i,j] = 1$ iff $\langle i,j \rangle \in f$, $\forall i \in \text{DOM}(f)$, $\forall j \in \text{CODOMAIN}(f)$

This yields matrices $\boldsymbol{S} = \boldsymbol{M}_{\text{S-SELECT}}$, $\boldsymbol{\Pi}_k = \boldsymbol{M}_{\text{PROJECTION}_K}$, and $\boldsymbol{\varepsilon}_k = \boldsymbol{M}_{\text{NOISE}_K}$. Figure 1 shows what these would look like for our running example.

We also need the analogue of function composition for the matrix forms $\boldsymbol{S}$ and $\boldsymbol{\Pi}_i$. This is given by the disjunctive normal form in (10).[4]

(10) $\boldsymbol{S}\boldsymbol{\Pi}_i[\text{VERB}, \text{FRAME}] = \bigvee_{\text{TYPE}} \boldsymbol{S}[\text{VERB}, \text{TYPE}] \wedge \boldsymbol{\Pi}_i[\text{TYPE}, \text{FRAME}]$

---

3 This formalization ignores the fact that, in practice, (7b) would either need to specify which semantic type signature the subcategorization frame is related to or give some general rule for determining it.

4 The definition in (8c) is just matrix multiplication for boolean algebras, which replaces sums with disjunctions and products with conjunctions.

$$S = \begin{array}{c} \\ \text{ask} \\ \text{wonder} \end{array} \begin{array}{cc} [\_\,Q] & [\_\,\text{Ent}] \\ \left(\begin{array}{cc} 1 & 0 \\ 1 & 0 \end{array} \right) \end{array}$$

$$\Pi_1 = \begin{array}{c} [\_\,Q] \\ [\_\,\text{Ent}] \end{array} \begin{array}{cc} [\_\,S] & [\_\,NP] \\ \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \end{array} \qquad \varepsilon_1 = \begin{array}{c} \text{ask} \\ \text{wonder} \end{array} \begin{array}{cc} [\_\,S] & [\_\,NP] \\ \left(\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right) \end{array}$$

$$\Pi_2 = \begin{array}{c} [\_\,Q] \\ [\_\,\text{Ent}] \end{array} \begin{array}{cc} [\_\,S] & [\_\,NP] \\ \left(\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right) \end{array} \qquad \varepsilon_2 = \begin{array}{c} \text{ask} \\ \text{wonder} \end{array} \begin{array}{cc} [\_\,S] & [\_\,NP] \\ \left(\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right) \end{array}$$

Figure 1 Boolean model for the running example

This allows us to define boolean matrices $D_1 \equiv S\Pi_1 \vee \varepsilon_1$ and $D_2 \equiv S\Pi_2 \wedge \varepsilon_2$ for DISTRIBUTION$_1$ and DISTRIBUTION$_2$. The only difference between $D_1 = D_2$ is how the information that is independent of S-selection and projection is handled.

We will eventually want to allow models to consider both of these possibilities at the same time—i.e., to learn whether particular bits of noise are additive or multiplicative. To this end, we generalize these approaches into a single approach that uses bitwise XOR (notated $\oplus$), which generates a 1 at a position iff exactly one of the inputs is 1. Instead of combining ε_k with $S\Pi_k$ via (bitwise) $\vee$ or $\wedge$, we instead specify an ε'_k in terms of ε_k that is combined via $\oplus$.

(11) a. $D_1 = S\Pi_1 \vee \varepsilon_1 = S\Pi_1 \oplus ((\varepsilon_1 \vee S\Pi_1) \oplus S\Pi_1) = S\Pi_1 \oplus \varepsilon'_1$
     b. $D_2 = S\Pi_2 \wedge \varepsilon_2 = S\Pi_2 \oplus ((\varepsilon_2 \wedge S\Pi_2) \oplus S\Pi_2) = S\Pi_2 \oplus \varepsilon'_2$

The 'bit-flipping' properties of $\oplus$ allow for an ε'_k that contains a 1 where $S\Pi_k$ makes the wrong prediction and a 0 wherever it makes the right one—i.e. a 1 now represents a case where the observed lexical distribution is different (in whichever way is relevant) from what the projection rules themselves would lead to. The transformed noise matrices for the running example are shown in Figure 2.

In statistical terminology, ε'_k can be thought of as a *matrix of residuals* or *error matrix* for $S\Pi_k$. The sum over ε'_k is then the *total error* or *loss* for $S\Pi_k$, abstracted over both additive and multiplicative lexical noise.

2.3 Upshots

The upshot of this translation and generalization of the additive and multiplicative approaches is two-fold. First, because we now have a way of representing S-selection and projection as matrices, we can take full advantage of algorithms that optimize over matrices. This is important because it allows us to construct inference

$$\boldsymbol{\varepsilon}'_1 = \begin{matrix} \\ \text{ask} \\ \text{wonder} \end{matrix} \begin{matrix} [\_\,S] & [\_\,NP] \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \end{matrix} \qquad \boldsymbol{\varepsilon}'_2 = \begin{matrix} \\ \text{ask} \\ \text{wonder} \end{matrix} \begin{matrix} [\_\,S] & [\_\,NP] \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix}$$

Figure 2   Transformed noise matrices for running example, for use with XOR.

algorithms that induce representations of matrices like S-SELECT, PROJECTION, and NOISE from data about verbs' observed syntactic distributions. A good chunk of Section 4 is devoted to showing how this is done.

Second, it means that, as long as we deal with lexical idiosyncrasy correctly—by ensuring that our representation of NOISE is an error matrix—we can remain agnostic about whether a Grimshaw- or Pesetsky-style system is correct. Taking advantage of this second upshot is highly dependent on the nature of the data at hand, and we will in fact need to further generalize the model beyond binary acceptability and XOR. Distributions representing selectional information are almost never actually representable in bits, even when we consider standard intuition-based judgment methods. That is, the claim that some set of verbs is grammatical with some set of surface syntactic frames involves the aggregation of *acceptability* judgments about particular sentences constructed from those frames. Acceptability itself likely must be treated, if not continuously, at least as ordinal (Chomsky 1965; Sprouse 2007b,a; White & Sprouse 2016). Before showing how to generalize from simple bitwise XOR to acceptability data, we turn next to the exact form of data we will be modeling in this paper: our large-scale acceptability judgment experiment.
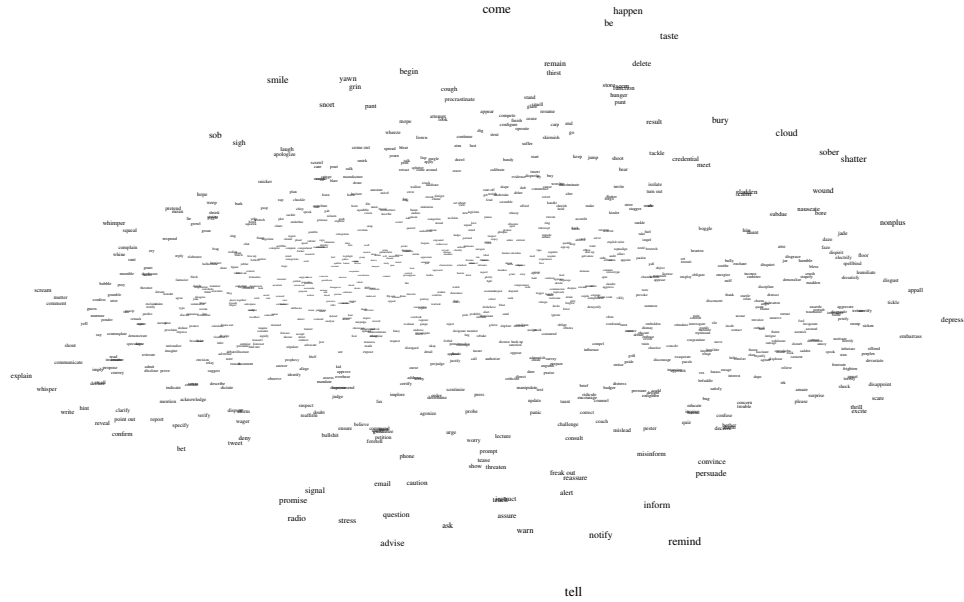
3   The MegaAttitude dataset

In this section, we report on the collection of data relevant to the syntactic distributions of essentially all clause-embedding verbs in English. By our count, there are on the order of 1,000 such verbs. The main obstacle to studying such a large number of verbs is obtaining the relevant data. One possible source might be syntactic distributions found in a corpus.[5] But the use of corpora—even very large corpora—introduces extra complications, such as genre effects (Roland & Jurafsky 1998) and data sparsity for low frequency verbs. Instead, we use a more direct measure of native speakers' competence: acceptability judgments.[6]

---

5 See Lederer, Gleitman & Gleitman 1995; Schulte im Walde 2000; Schulte im Walde & Brew 2002; Schulte im Walde 2006, 2003; Korhonen 2002; Buttery & Korhonen 2005; Buttery 2006; Alishahi & Stevenson 2008; Barak, Fazly & Stevenson 2012, 2013, 2014a,b; White 2015 among others.

6 The model we present in Section 4 does not require access to acceptability judgment data (see White, Hacquard & Lidz under review for preliminary results on corpus data). Indeed, it could in principle be applied to any type of data for which an appropriate link function can be designed. See, e.g.,

Figure 3   1,000 verbs in the MegaAttitude data set.

3.1   Methodology and materials

To gather this more direct measure, we adopt a methodology introduced by Fisher, Gleitman & Gleitman (1991) and adapted by White, Dudley, Hacquard & Lidz (2014) to the study of clause-embedding verbs; here we adapt this method to a very large data set.[7]   In this methodology, one (i) selects a set of verbs and a set of subcategorization frames, (ii) computes the cartesian product of these sets, (iii) constructs sentences for each element of this cartesian product, and (iv) gathers acceptability judgments about those sentences on an ordinal (1-7) scale.

We selected 1,000 clause-embedding verbs to include in our large-scale experiment using a search of VerbNet (Kipper-Schuler 2005) seeded by the clause-embedding verbs discussed in Hacquard & Wellwood 2012; Anand & Hacquard 2013, 2014; Rawlins 2013; White et al. 2014. This seed contained ∼500 verbs. Figure 3 shows all 1,000 verbs.[8]

White's (2015) proposed negative binomial link, which is directly portable to our model.

7 We validated the resulting methodology in three pilots. We do not have space here to describe these pilots in detail, but we report on them in a longer paper that is in preparation.

8 In the electronic version, this image is high-enough resolution to zoom in quite close. The supplementary materials include a much larger version of the same figure.
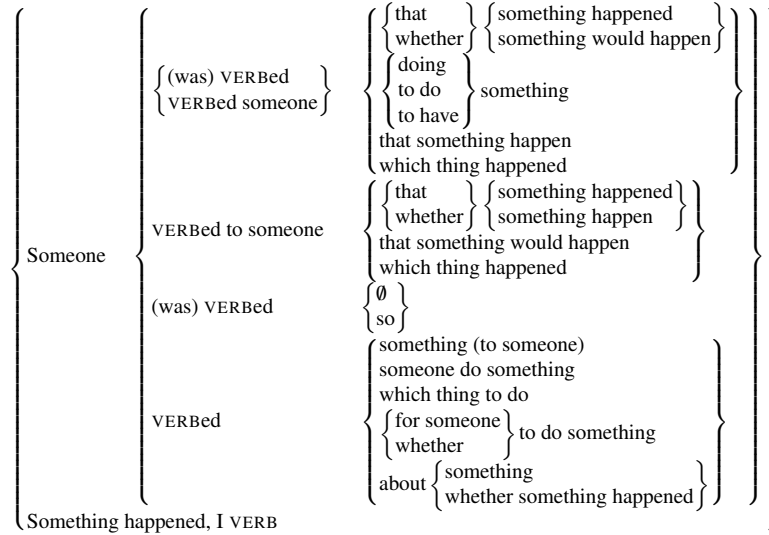
Figure 4   50 frames constructed for the MegaAttitude experiment.

We constructed 50 subcategorization frames representing different combinations of syntactic features purported to be relevant to clause-embedding verb semantics (see White 2015 for a review). Six basic types of syntactic features were employed, listed in (12). Figure 4 shows the 50 frames constructed using these features.[9]

(12)   a.   COMPLEMENTIZER: *that*, *for*, ∅, *whether*, *D-linked WH*
        b.   EMBEDDED TENSE: *past*, *future*, *infinitival*, *present participle*, *bare*
        c.   MATRIX OBJECT: *none*, *one*, *two*
        d.   MATRIX PP: *none*, *to*, *about*
        e.   EMBEDDED SUBJECT: *true*, *false*
        f.   PASSIVIZATION: *true*, *false*

We constructed sentences from verb-frame pairs in a way similar to Fisher et al. 1991 and White et al. 2014, but employing the following constraint: all lexical

---

9 Passivized sentences, which all had active counterparts, were included to capture verbs that take expletive subjects. This was done instead of including explicit expletive subject frames because such frames are systematically ambiguous between expletive and referential readings, and thus they do not pick out solely the expletive subject-taking verbs but rather the somewhat unnatural class of expletive subject-taking verbs plus referential subject verbs that can take an inanimate subject (see White et al. 2014; White, Hacquard & Lidz under revision; White 2015 for discussion). Using passivization provides a cleaner cut in that expletive subject frame are exactly those verbs that are good in the passive but bad in the active, though at the cost of missing verbs that cannot passivize—e.g., verbs of relevance (Karttunen 1977a).
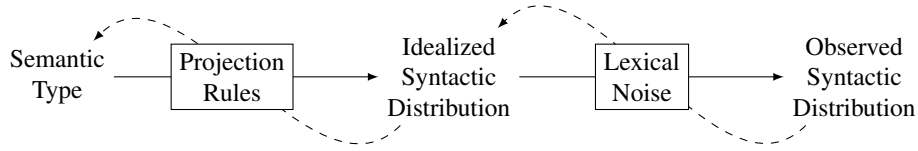
Figure 5 Relationships between various model components (solid arrows) and schematization of inference (dashed arrows)

categories in the sentence besides the verb should be as bleached as possible. This constraint is intended to avoid typicality effects induced by, e.g., sentence plausibility (see Schütze & Sprouse 2014). We deployed this constraint by instantiating all NPs as either *someone* or *something*, all untensed VPs as either *do something* or *have something*, and all Ss as COMP+*something happen*+TENSE/MODALITY. Figure 4 shows the full set of frames. The 50,000 resulting sentences were partitioned into 1,000 lists of 50 such that each verb and each frame occur at most once in each list.

3.2 Participants

727 unique participants were recruited through Amazon Mechanical Turk to rate sentences in the 1,000 lists of 50. Participants were allowed to respond to as many unique lists as they liked, but no one participant rated the same list twice. Five ratings were obtained per sentence—i.e., five unique participants rated each list—and each participant responded to a median of four lists (mean: 6.9, min: 1, max: 56). The end result is a data set with 250,000 ordinal acceptability judgments.

These data were then filtered to remove data from participants that showed poor agreement with other participants. A total of 27 participants were removed in this way.[10] The final filtered data set had 35,642 items with five judgments, 12,506 items with four judgments, and 1,852 items with three judgments.

4 Computational model

In this section, we develop our computational model of S-selection based on the formalization laid out in Section 2. This formalization has five components, whose relationships are schematized in Figure 5 as solid arrows. The only thing the model gets to learn from are observed syntactic distributions—in the current case, the

---

10 Space does not allow for a full explication of this somewhat complex filtering procedure, which takes into account factors such as the list each item was found in, the verb that the item instantiates, the frame that the item instantiates, and the item itself.

acceptability judgment data described in the last section. The goal is to use these data and the aforementioned relationships to (i) infer back through the lexical noise to the idealized syntactic distributions and (ii) infer back through the projection rules to the semantic type representations. In what follows we outline the main components of inference. There are four key parts: (i) a probabilistic version of the boolean formalization with the structure shown in Figure 5, (ii) a way of linking this probabilistic model to the ordinal data, (iii) optimization procedures for training the model, and (iv) a model selection procedure to find the best model.

4.1 A probabilistic model

To conduct the inference schematized in Figure 5, we must specify some method by which search the space of possible representations for an optimal one. One possibility is to take exactly the model specified in Section 2 and attempt to find boolean matrix representations of S-selection and projection that minimize the amount of lexical noise we have to posit. There are two issues with this approach. The first is that this optimization over boolean matrices is intractable due to the sheer number of possibilities: with V verbs, T semantic type signatures, and F frames, there are $\left|\{0,1\}^{V \times T} \times \{0,1\}^{T \times F}\right| = 2^{VT+TF}$ configurations to search over. The second issue is that encoding ordinal judgments as boolean values loses potentially useful information about relative acceptability in the ordinal judgments. To address these issues, we develop a probabilistic version of the formalization in Section 2.

As a reminder, this formalization consists of four interrelated matrix representations: the S-selection matrix $\boldsymbol{S} \in \{0,1\}^{V \times T}$; the projection matrix $\boldsymbol{\Pi} \in \{0,1\}^{T \times F}$; the error/noise matrix $\boldsymbol{\varepsilon} \in \{0,1\}^{V \times F}$; and the syntactic distribution matrix $\boldsymbol{D} \in \{0,1\}^{V \times F}$. $\boldsymbol{D}$ is deterministically related to $\boldsymbol{S}$ and $\boldsymbol{\Pi}$ by $\boldsymbol{D} \equiv \boldsymbol{S}\boldsymbol{\Pi} \oplus \boldsymbol{\varepsilon}$.

Going forward, it is useful to have a variable name for the idealized syntactic distribution matrix $\boldsymbol{S}\boldsymbol{\Pi}$. We refer to this matrix as $\hat{\boldsymbol{D}}$.

(13) $\hat{\boldsymbol{D}}[v,f] = \boldsymbol{S}\boldsymbol{\Pi}[v,f] = \bigvee_t \boldsymbol{S}[v,t] \wedge \boldsymbol{\Pi}[t,f]$

We convert this formalization into a probabilistic model by 'wrapping' each matrix element with a probability function, making the following independence assumptions: $\mathbb{P}(\boldsymbol{S},\boldsymbol{\Pi}) = \mathbb{P}(\boldsymbol{S})\mathbb{P}(\boldsymbol{\Pi})$ and $\mathbb{P}(\boldsymbol{S})\mathbb{P}(\boldsymbol{\Pi}) = \prod_t \prod_v \mathbb{P}(\boldsymbol{S}[v,t]) \prod_f \mathbb{P}(\boldsymbol{\Pi}[t,f])$. The first of these amounts to the assumption that projection rules are general to all verbs. The second amounts to a strong assumption about the relationship among type signatures—namely, that one's behavior carries no information about the others'.[11] Together, they allow us to straightforwardly calculate $\mathbb{P}(\hat{\boldsymbol{D}}[v,f])$.

---

11 The second assumption is almost certainly too strong and is related to (though not a consequence of) the fact that we do not represent type signatures as structured objects. We return to this in Section 6.

(14) $\mathbb{P}\left(\hat{\boldsymbol{D}}[v,f]\right) = 1 - \prod_t 1 - \mathbb{P}(\boldsymbol{S}[v,t])\mathbb{P}(\boldsymbol{\Pi}[t,f])$

Thus, given $\mathbb{P}(\boldsymbol{S})$ and $\mathbb{P}(\boldsymbol{\Pi})$, we can analytically compute $\mathbb{P}(\hat{\boldsymbol{D}}[v,f])$. But how do we similarly compute $\mathbb{P}(\boldsymbol{D}[v,f])$? We could play the same game we did before by assuming that $\forall v, f : \mathbb{P}(\hat{\boldsymbol{D}}[v,f], \boldsymbol{\varepsilon}[v,f]) = \mathbb{P}(\hat{\boldsymbol{D}}[v,f])\mathbb{P}(\boldsymbol{\varepsilon}[v,f])$, thus yielding (15).

(15) $\mathbb{P}\left(\boldsymbol{D}[v,f]\right) = \mathbb{P}\left(\hat{\boldsymbol{D}}[v,f]\right)(1 - \mathbb{P}\left(\boldsymbol{\varepsilon}[v,f]\right)) + \left(1 - \mathbb{P}\left(\hat{\boldsymbol{D}}[v,f]\right)\right)\mathbb{P}\left(\boldsymbol{\varepsilon}[v,f]\right)$

The problem is that this probabilistic representation does not straightforwardly map to the kind of data we have: it assumes some binary representation, where we have an ordinal representation. We thus require some generalized version of (15).

4.2 Ordinal link function

An ordinal link function can be thought of as a generalized version of (15). Without getting into the fine details of ordinal links, the basic idea is to cut up some range of (usually) continuous values—in our case, $\hat{\boldsymbol{D}}[v,f] \in [0,1]$—into contiguous subintervals labeled in ascending order by the ordinal scale points. One then constructs a function from the range of continuous values into probability distributions over the scale points that uses the collection of subintervals. The particular kind of ordinal link we use allows for the collection of subintervals that determines the link function to vary with each participant. That is, each participant may cut up the $[0,1]$ interval in a slightly different way.[12]

4.3 Model fitting

For a fixed number of semantic type signatures T, we use projected gradient descent with adaptive gradient (Duchi, Hazan & Singer 2011) to optimize $\boldsymbol{S} \in [0,1]^{V \times T}$ and $\boldsymbol{\Pi} \in [0,1]^{T \times F}$ to maximize the likelihood of the acceptability data. Thus, for each setting of T we obtain representations $\boldsymbol{S}_{\text{best}(T)} \in [0,1]^{V \times T}$ and $\boldsymbol{\Pi}_{\text{best}(T)} \in [0,1]^{T \times F}$ that are optimal in the sense that they explain the data as well or better than any other $\boldsymbol{S} \in [0,1]^{V \times T}$ and $\boldsymbol{\Pi} \in [0,1]^{T \times F}$.

4.4 Model comparison and selection

One thing we cannot tell from the above model fitting procedure alone is what the 'right' value of T is. In fact, ideally, we could analyze what the model does at many different values of T. But for reasons of space, this is not feasible. Therefore, we need some way of selecting a model to look at.

---

12 See White et al. under revision for evidence that such an ordinal link is optimal for data collected under Fisher et al.'s (1991) paradigm

The problem of selecting such a value is common in the statistics and machine learning literature. One standard method is to fit the model with various settings of T, then calculate both how well the model explains the data—i.e., how small the total error is—and how much explanatory power the model had to work with while fitting the data. The best model is the one that maximizes explanation while minimizing explanatory power. To carry out such a selection procedure, one needs a way of quantifying the trade-off between fit and explanatory power.

We use the Akaike Information Criterion (AIC) to do this (Akaike 1974). The AIC measures the information-theoretic 'distance' between a model and the true (but unknown) data-generating process using the Kullback-Liebler divergence. The optimal model under AIC is the one that minimizes this distance.

We find that a model with 12 semantic type signatures minimizes the AIC across all values of T. The way to interpret this result is that 12 is a lower bound on the number of semantic type signatures we should posit for the space of clause-embedding verbs. The extent to which 12 is a good upper bound is dependent on (i) how exhaustively we were able to sample the full space of clause-embedding verbs; (ii) how exhaustively we were able to sample the full space of semantically relevant subcategorization frames, and (iii) whether all possible semantic type signatures relevant to clause-embedding are actually instantiated in English.

The full projection matrix for these 12 type signatures is shown in Figure 6, and two slices of the S-selection matrix are shown in Figures 8 and 9. We analyze these in the next section. In the remainder of the paper, we discuss how to interpret these type signatures, focusing on a case study of interrogative v. declarative clauses.

5 Results

In this section, we present the results of the fitting and selection procedure, analyzing only a subset of these results for reasons of space. (We provide a full analysis in a paper in preparation.) As a case study we focus on semantic type signatures relevant to declarative and interrogative selection. We noted in Section 1 that explaining the distribution of declarative and interrogative embedded clauses is a major question in the literature on clause-embedding verbs, and so we begin with a very brief overview of the proposals that have been instantiated in this literature. Based on our results, we then argue that many of these proposals are correct to some extent, though only when relativized to particular verb classes.

5.1 Embedding declaratives and interrogatives

There are three types of verbs that take either declaratives or interrogatives: those like *believe*, which take only declaratives; those like *wonder*, which take only
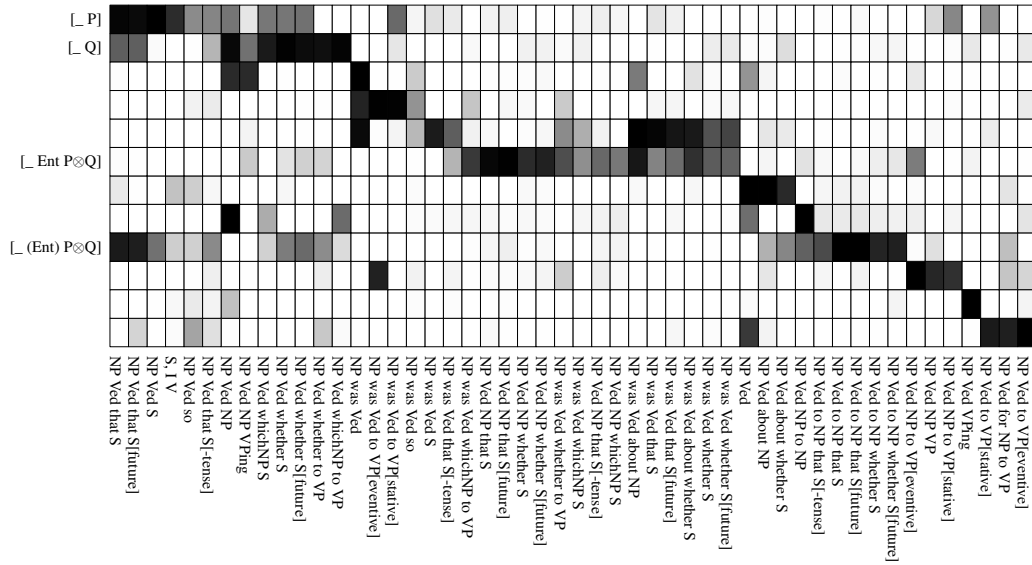
Figure 6 Representation of projection rules for all semantic type signatures. (Only labeled rows are discussed in the present paper.)

interrogatives; and those like *know*, which take both (Lahiri's 2002 *responsives*).

(16)   a.   John believed {that, *whether} Mary went to the store.
       b.   John wondered {*that, whether} Mary went to the store.
       c.   John knew {that, whether} Mary went to the store.

Responsives have proven difficult to capture. There are roughly three approaches in the literature to explaining the syntactic their distribution: (i) responsives are ambiguous between a declarative- and an interrogative-taking variant—e.g., they S-select both propositions and questions (Karttunen 1977a; Heim 1994; George 2011);[13] (ii) responsives only S-select propositions, and interrogatives either (iia) always denote propositions (Groenendijk & Stokhof 1984), or (iib) they are somehow cast into propositions by an operator or coercion rule (Egré 2008; Spector & Egre 2015); or (iii) responsives only S-select questions, and declaratives (iiia) either always denote questions (Rawlins 2013), or (iiib) they are somehow cast into questions by an operator or rule (Uegaki 2012, 2015a,b).

We cannot do justice to this rich literature here, but even at this coarse level of analysis we can fruitfully use the abstract characterizations above to make predictions about the results our model should produce if any one option is correct. In particular, option (i) and options (ii) and (iii) pull apart in the following way: option (i) is at

---

13 We include approaches that posit regular polysemy under this heading (cf. George 2011).
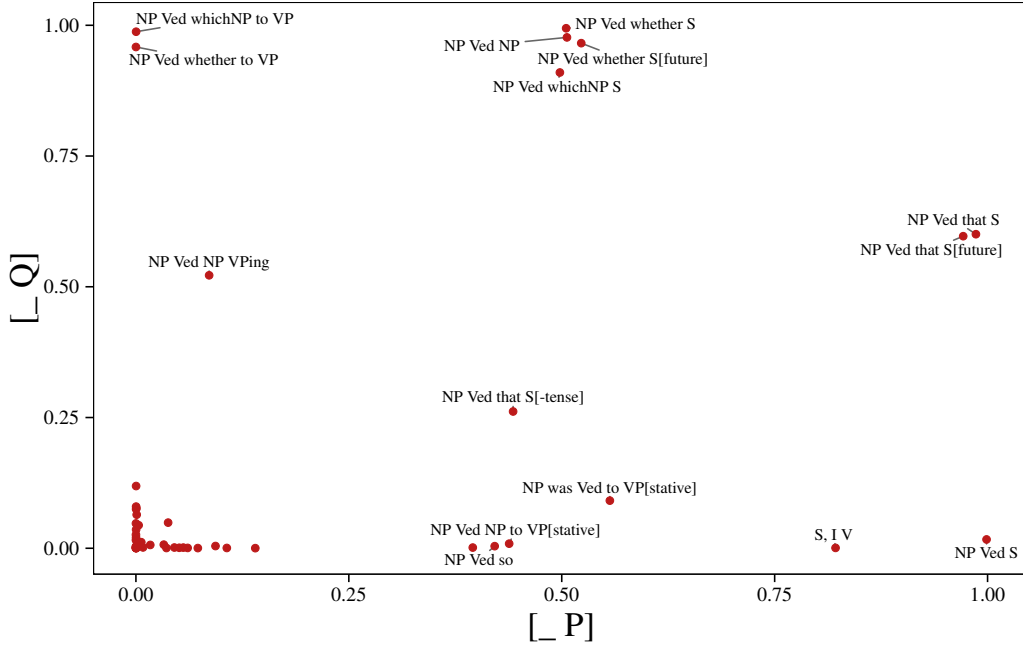
**Figure 7**   Probability of projection rules for propositions and questions

least compatible with there being two distinct projection rules for responsives—one that projects onto declaratives and another that projects onto questions; in contrast, options (ii) and (iii) should posit only a single such rule that projects onto both.

What we in fact find is that both kinds of projection rules exist, but not for the same verb classes. Verbs that S-select a recipient—i.e., communicative verbs like *say* and *tell*—S-select a single type signature that projects onto both interrogatives and declaratives, while verbs that do not S-select a recipient—e.g., cognitive verbs like *know*—S-select two distinct type signatures: one that projects onto declaratives and another that projects onto interrogatives. Furthermore, in the second case, we find quantitative evidence that some kind of coercion can operate in both directions.

5.2   Cognitives v. communicatives

Figure 6 depicts the probabilistic projection rules our model learns. Along the columns lie the full set of subcategorization frames we tested in our experiment and along the rows are the 12 semantic type signatures our model infers (see Section 4.4). The model itself does not label these type signatures; we apply labels in Figure 6 based on the subcategorization frames they project onto, and we have labeled only those type-signatures we discuss in detail in this paper.
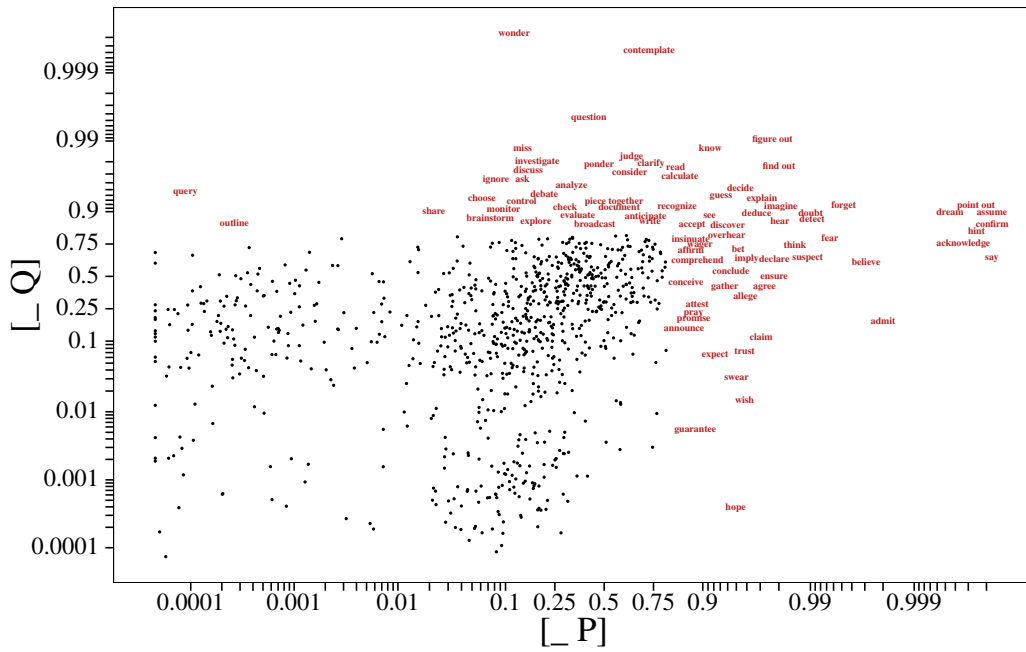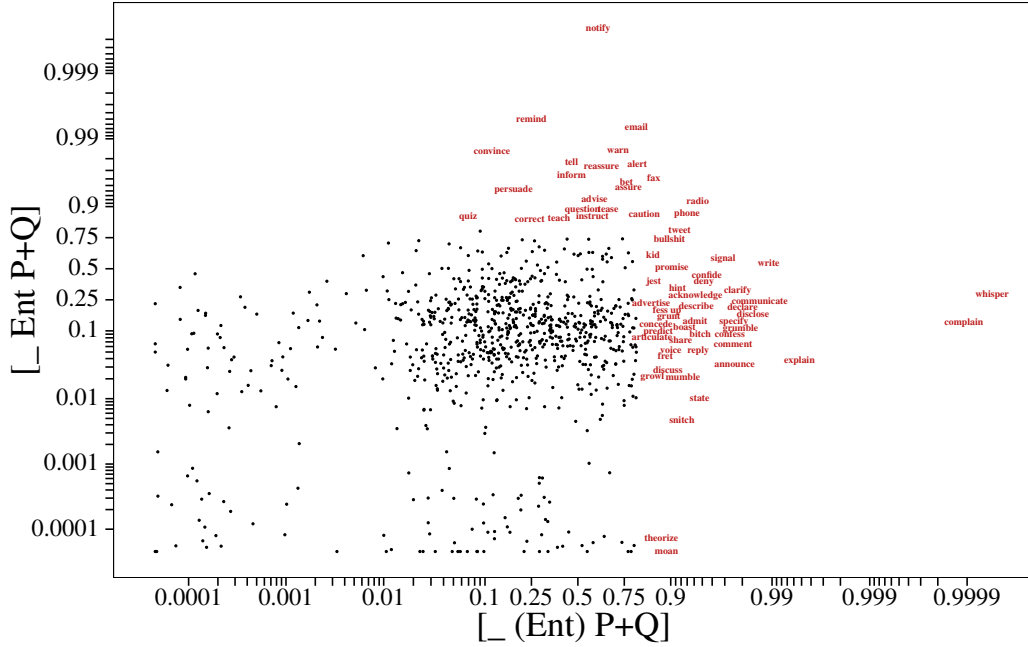
Figure 8   Probability of S-selection for propositions and questions

There are three things to note about Figure 6. First, there are two distinct type signatures projecting onto declaratives and interrogatives when the clause is not accompanied by another internal argument. We label these type signatures [_ P] and [_ Q], respectively. Second, when the clause is accompanied by another internal argument—e.g., an NP or PP—there is a single type signature that projects onto both declaratives and interrogatives: one that projects onto subcategorization frames with an NP internal argument and another that projects onto subcategorization frames with a PP internal argument. We label these [_ Ent P⊗Q] and [_ (Ent) P⊗Q], respectively, and we refer to them as *hybrid* types. Finally, we see that, though [_ P] and [_ Q] project most strongly onto declaratives and interrogatives, there is middling probability that [_ P] projects onto interrogatives and [_ Q] onto declaratives.

This last result can be seen more clearly in Figure 7, which plots the probability that [_ P] and [_ Q] project onto each frame. Importantly, we see that no frames are strongly probable under both [_ P] and [_ Q]—i.e., there are no frames in the upper right corner—and there are frames that are strongly probable for [_ P] (rightmost) and strongly probable for [_ Q] (topmost). The frames with middling probability for [_ Q] are exactly the declaratives with a complementizer, and the frames with middling probability for [_ P] are the finite interrogatives (plus the NP direct object frame). We conclude from this that our model is detecting coercion of a question

Figure 9   Probability of S-selection for hybrid types with recipients

into a proposition and coercion of a proposition into a question.

The verbs that take the [_ P] and [_ Q] type signatures are plotted in Figure 8. We see that, as one would expect if these were in fact the type signatures [_ P] and [_ Q], *wonder* takes only [_ Q] with high probability, *believe* takes only [_ P] with high probability, and *know* takes both [_ P] and [_ Q] with high probability.

The verbs that take the [_ (Ent) P⊗Q] and [_ Ent P⊗Q] type signatures are plotted in Figure 9. We see that this set includes only communicative verbs, suggesting that the entity in this type signature has some sort of RECIPIENT role. One reason that communicatives might have hybrid types is that they may encode reference to Question Under Discussion as a part of their semantics (Anand & Hacquard 2014). This in turn may be thought of as requiring a hybrid type that can represent multiple kinds of content—e.g., informative and inquisitive content (cf. Groenendijk 2009; Ciardelli, Groenendijk & Roelofsen 2013; Rawlins 2013).

In summary, for cognitive verbs, our modeling results support some version of option (i), where there are both proposition- and question-selecting type signatures, combined with coercion. For communication verbs, the results support instead an approach where the verb selects just one hybrid type. While we cannot definitively rule out the coercion approaches, these results are suggestive of an approach where there is no coercion at the level of semantic types for communicatives.

6   Conclusion

In this paper, we developed a computational model of S-selection that encodes both the notion of systematic mappings from semantic type signature to syntactic distribution—i.e., projection rules—and the notion of selectional noise—e.g., C(ategory)-selection, L(exical)-selection, and/or other independent syntactic processes. We then collected a large-scale data set assessing the acceptability of 1,000 English clause-embedding verbs in 50 distinct syntactic frames and fit the model to these data. Finally, we analyzed a subset of the results, relating them to a long-standing in the literature on clause-embedding verbs and arguing that previous accounts need to be relativized to particular verb classes.

We intend this model to be a first step in the longer term development of theoretically informed computational models. In this last section, we conclude with some ways that one could further develop both the model itself as well as analyses of its output to ask finer-grained theoretical questions.

A straightforward place to start is by asking to what extent our results favor an additive (Grimshaw-style) or multiplicative (Pesetsky-style) account of selection. Our model abstracts across these approaches, but in principle the errors it makes can be transformed into a purely additive or purely multiplicative representation. This would allow us to ask, e.g., which approach requires less lexical idiosyncrasy.

Another area to further develop is how the model represents the semantic type signatures it does inference over. Currently, the model has no access *a priori* to what the space of semantic type signatures is or to any relationships among those type signatures. That is, nowhere does it encode that semantic type signatures such as [_ Q], [_ Ent], [_ Ent Q] are possible or that they share constituent types. The same is true of the subcategorization frames.

In current work, we have been developing ways of inferring the structure of the semantic type representations by explicitly representing (i) type constructors for mapping raw types to structured type signatures; (ii) relationships between verbs and raw types; and (iii) relationships between raw types and syntactic constituents. We believe that, beyond providing further insight into the structure of the type signature space our model learns, these developments may provide insight into fundamental questions about regular polysemy in the clause-embedding lexicon.

The main idea here is that, if at least some forms of regular polysemy manifest themselves as distinctions in semantic type signature (Pustejovsky 1991, 1995; Rappaport Hovav & Levin 1998, 2001), we should in principle be able to apply to our model's semantic type signatures the same inference algorithm we applied to syntactic distributions. This is essentially what the more elaborated model just described does. Preliminary investigations of this elaborated model provide initial evidence for regular polysemy among cognitive responsives (cf. George 2011).

References

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

Alishahi, Afra & Suzanne Stevenson. 2008. A computational model of early argument structure acquisition. *Cognitive Science* 32(5). 789–834.

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

Anand, Pranav & Valentine Hacquard. 2013. Epistemics and attitudes. *Semantics and Pragmatics* 6(8). 1–59.

Anand, Pranav & Valentine Hacquard. 2014. Factivity, belief and discourse. In Luka Crnič & Uli Sauerland (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim* 1, 69–90. Cambride, MA: MIT Working Papers in Linguistics.

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

Barak, Libby, Afsaneh Fazly & Suzanne Stevenson. 2012. Modeling the acquisition of mental state verbs. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, 1–10. Association for Computational Linguistics.

Barak, Libby, Afsaneh Fazly & Suzanne Stevenson. 2013. Acquisition of Desires before Beliefs: A Computational Investigation. In *Proceedings of the 17th Conference on Computational Natural Language Learning (CoNLL)*, .

Barak, Libby, Afsaneh Fazly & Suzanne Stevenson. 2014a. Gradual Acquisition of Mental State Meaning: A Computational Investigation. In *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society*, .

Barak, Libby, Afsaneh Fazly & Suzanne Stevenson. 2014b. Learning verb classes in an incremental model. *ACL 2014* 37.

Bresnan, Joan W. 1972. *Theory of complementation in English syntax.*: Massachusetts Institute of Technology PhD dissertation.

Buttery, Paula. 2006. *Computational models for first language acquisition*: University of Cambridge PhD dissertation.

Buttery, Paula & Anna Korhonen. 2005. Large-scale analysis of verb subcategorization differences between child directed speech and adult speech. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, .

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

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* 11. Cambridge, MA: MIT press.

Chomsky, Noam. 1973. Conditions on transformations. In S. Anderson & P. Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart, & Winston.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2013. Inquisitive semantics: a new notion of meaning. *Language and linguistic compass* 7(9). 459–476.

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

Egré, Paul. 2008. Question-embedding and factivity. *Grazer Philosophische Studien* 77(1). 85–125.

Fillmore, Charles John. 1970. The Grammar of Hitting and Breaking. In R.A. Jacobs & P.S. Rosenbaum (eds.), *Readings in English Transformational Grammar*, 120–133. Waltham, MA,: Ginn.

Fisher, Cynthia, Henry Gleitman & Lila R. Gleitman. 1991. On the semantic content of subcategorization frames. *Cognitive psychology* 23(3). 331–392.

Frana, Ilaria. 2010. *Concealed Questions. In Search of Answers.*: University of Massachusetts, Amherst PhD dissertation.

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

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

Ginzburg, Jonathan & Ivan Sag. 2001. *Interrogative investigations*. Stanford: CSLI publications.

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

Grimshaw, Jane. 1981. Form, function and the language acquisition device. In *The Logical Problem of Language Acquisition*, 165–182.

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

Groenendijk, Jeroen. 2009. Inquisitive semantics: Two possibilities for disjunction. In P. Bosch, D. Gabelaia & J. Lang (eds.), *Seventh International Tbilisi Symposium on Language, Logic, and Computation*, Springer.

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

Gruber, Jeffrey Steven. 1965. *Studies in lexical relations*: Massachusetts Institute of Technology PhD 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 *Semantics from different points of view*, 51–60. Springer.

Heim, Irene. 1994. Interrogative semantics and Karttunen's semantics for know. In

Proceedings of IATL 1, 128–144.

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

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

Karttunen, Lauri. 1977b. To doubt whether. In *The CLS Book of Squibs*, Chicago: Chicago Linguistic Society.

Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*: University of Pennsylvania PhD dissertation.

Korhonen, Anna. 2002. *Subcategorization Acquisition*: University of Cambridge PhD dissertation.

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

Lederer, Anne, Henry Gleitman & Lila Gleitman. 1995. Verbs of a feather flock together: Semantic information in the structure of maternal speech. *Beyond Names for Things: Young children's acquisition of verbs* 277.

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

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

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

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

Pesetsky, David. 1995. *Zero syntax: Experiencers and cascades*. MIT press.

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.

Pustejovsky, James. 1991. The generative lexicon. *Computational linguistics* 17(4). 409–441.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press. https://mitpress.mit.edu/books/generative-lexicon.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. *The projection of arguments: Lexical and compositional factors* 97–134.

Rappaport Hovav, Malka & Beth Levin. 2001. An event structure account of English resultatives. *Language* 766–797.

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

Roland, Douglas & Daniel Jurafsky. 1998. How verb subcategorization frequencies are affected by corpus choice. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on*

Computational Linguistics-Volume 2, 1122–1128. Association for Computational Linguistics.

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

Schütze, Carson T. & Jon Sprouse. 2014. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 27–50. Cambridge University Press.

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

Sprouse, Jon. 2007a. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1. 123–134.

Sprouse, Jon. 2007b. *A program for experimental syntax*: University of Maryland PhD dissertation.

Stone, Marshall H. 1936. The theory of representation for Boolean algebras. *Transactions of the American Mathematical Society* 40(1). 37–111.

Uegaki, Wataru. 2012. Content nouns and the semantics of question-embedding predicates. *Proceedings of Sinn und Bedeutung 16* .

Uegaki, Wataru. 2015a. Content nouns and the semantics of question-embedding. *Journal of Semantics* .

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

Schulte im Walde, Sabine. 2000. Clustering verbs semantically according to their alternation behaviour. In *Proceedings of the 18th conference on Computational linguistics-Volume 2*, 747–753. Association for Computational Linguistics.

Schulte im Walde, Sabine. 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*: Universität Stuttgart PhD dissertation.

Schulte im Walde, Sabine. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics* 32(2). 159–194.

Schulte im Walde, Sabine & Chris Brew. 2002. Inducing German semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 223–230. Association for Computational Linguistics.

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

White, Aaron Steven, Rachel Dudley, Valentine Hacquard & Jeffrey Lidz. 2014. Discovering classes of attitude verbs using subcategorization frame distributions. In Hsin-Lun Huang, Ethan Poole & Amanda Rysling (eds.), *Proceedings of the 43rd annual meeting of the North East Linguistic Society*, .

White, Aaron Steven, Valentine Hacquard & Jeffrey Lidz. under review. Main clause syntax and the labeling problem in syntactic bootstrapping. In Kristen Syrett

& Sudha Arunachalam (eds.), *Semantics in Acquisition* Trends in Language Acquisition Research (TiLAR), John Benjamins Publishing Company.

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

White, Aaron Steven & Jon Sprouse. 2016. The trace of categorical structure in gradient judgments. Poster presented at the 34th West Coast Conference on Formal Linguistics, Salt Lake City, UT.

Zwicky, Arnold M. 1971. In a manner of speaking. *Linguistic Inquiry* 2(2). 223–233.

Aaron Steven White
Cognitive Science Department
Johns Hopkins University
3400 N Charles St.
Baltimore, MD, 21218
aswhite@jhu.edu

Kyle Rawlins
Cognitive Science Department
Johns Hopkins University
3400 N Charles St.
Baltimore, MD, 21218
kgr@jhu.edu