The labeling problem in syntactic bootstrapping
Main clause syntax in the acquisition of propositional attitude verbs*

Aaron Steven White
Johns Hopkins University

Valentine Hacquard
University of Maryland

Jeffrey Lidz
University of Maryland

Abstract

In English, the distinction between belief verbs, such as *think*, and desire verbs, such as *want*, is tracked by the tense of those verbs' subordinate clauses. This suggests that subordinate clause tense might be a useful cue for learning these verbs' meanings via *syntactic bootstrapping*. However, the correlation between tense and the belief v. desire distinction is not cross-linguistically robust; yet these verbs' acquisition profile is similar cross-linguistically. Our proposal in this chapter is that, instead of using concrete cues like subordinate clause tense, learners may utilize more abstract syntactic cues that must be tuned to the syntactic distinctions present in a particular language. We present computational modeling evidence supporting the viability of this proposal.

1 Introduction

Syntactic bootstrapping encompasses a family of approaches to verb learning wherein learners use the syntactic contexts a verb is found in to infer its meaning (Landau and Gleitman, 1985; Gleitman, 1990). Any such approach must solve two problems. First, it must specify how learners *cluster* verbs—i.e. figure out that some set of verbs shares some meaning component—according to their syntactic distributions. For instance, an English learner might cluster verbs based on whether they embed tensed subordinate clauses (1) or whether they take a noun phrase (2).

(1) a. John {thinks, believes, knows} that Mary is happy.
    b. *John {wants, needs, orders} that Mary is happy.

(2) a. John {believes, knows, wants, needs} Mary.
    b. *John {thinks, orders} Mary.

For different parts of the lexicon, different clusterings may be better or worse. Among propositional attitude verbs—like *think*, *believe*, *know*, *want*, *need*, and *order*—clustering based on whether the verb takes a subordinate clause yields intuitively better clusters than clustering based on whether it takes a noun phrase (at least when these structures are considered in isolation). That is, CLUSTERs 1 and 2 are intuitively more coherent than CLUSTERs 3 and 4 (see White et al. accepted for empirical corroboration of this intuition).

(3) a. CLUSTER 1: think, believe, know

*We are grateful to audiences at Semantics and Linguistic Theory 24, the University of Maryland, and UC Santa Cruz as well as Alexander Williams, Shevaun Lewis, Kaitlyn Harrigan, Naho Orita, Rachel Dudley, Morgan Moyer, and Tom Roberts for useful comments. This work was supported by NSF BCS DDRIG grant 1456013, NSF BCS grant 1124338, and NSF DGE IGERT grant 0801465. This is a preprint of a paper that is to appear in *Semantics in Acquisition*, Trends in Language Acquisition Research (eds. K. Syrett and S. Arunachalam). Amsterdam: John Benjamins Publishing Company. Comments are welcome and should be addressed to Aaron Steven White at aswhite@jhu.edu.

b. CLUSTER 2: want, need, order
c. CLUSTER 3: believe, know, want, need
d. CLUSTER 4: think, order

We refer to the problem of choosing how to cluster verbs based on syntactic context as the *clustering problem*, and we call the learning mechanism that solves this problem—i.e. outputs clusters like those in (3)—the *clustering mechanism*.

The second problem a syntactic bootstrapping approach must solve involves the method by which learners *label* the clusters output by a clustering mechanism—i.e. figure out what meaning component a particular cluster of verbs corresponds to. For instance, a common way of labeling CLUSTERs 1 and 2 is to say that all verbs in CLUSTER 1 have a BELIEF component and all verbs in CLUSTER 2 have a DESIRE component.

(4) a. CLUSTER 1 $\longleftrightarrow$ BELIEF
b. CLUSTER 2 $\longleftrightarrow$ DESIRE

We refer to this second problem, which is in many ways more difficult than the clustering problem, as the *labeling problem*, and we call the learning mechanism that solves this problem—i.e. labels the clusters output by the clustering mechanism—the *labeling mechanism*.

In this paper, we present evidence from the domain of propositional attitude verbs that previous labeling mechanisms are unsatisfactory both empirically and explanatorily, and we propose a novel labeling mechanism. Propositional attitude verbs are a useful case study for making this point because they are a parade case of verbs that fall prey to the *observability problem*—one cannot see propositional attitudes such as thinking or wanting—and thus likely require learners to rely heavily on syntactic evidence for their acquisition (Gleitman, 1990; Gillette et al., 1999; Snedeker and Gleitman, 2004; Gleitman et al., 2005; Papafragou et al., 2007).

We focus in particular on the distinction among propositional attitude verbs between belief verbs, like *think*, and desire verbs, like *want*, because properties of this distinction make the empirical and explanatory inadequacies of prior approaches particularly apparent. But our proposal has implications well beyond explaining this distinction—indeed, implications for the fundamental architecture of the learning mechanism itself.

In Section 2, we discuss the two main approaches to solving the labeling problem that have been instantiated in the literature—the *top-down approach* and the *bottom-up approach*—and show that both of these approaches are inadequate: the top-down approach makes incorrect predictions and the bottom-up approach makes essentially no predictions at all. In Section 3, we propose a modification to the top-down approach that makes correct predictions. In Section 4, we present a learning algorithm that implements our proposal. In Section 5, we present a proof-of-concept experiment, which shows that our algorithm finds a correct labeling when run on syntactic distributions found in child-direct speech (CDS). In Section 6, we discuss what our proposal entails for the theory of verb learning, and in Section 7, we conclude.

2 Approaches to the labeling problem

Current approaches to the labeling problem fall into two broad categories: the *top-down approach* and the *bottom-up approach*. In the top-down approach—the traditional one laid out in Landau and Gleitman 1985; Gleitman 1990—labeling is part-and-parcel with clustering. The learner has some innate mappings—*projection rules*—from semantic features to syntactic features (Gruber, 1965; Carter, 1976; Chomsky, 1981; Pinker, 1984, 1989; Grimshaw, 1990; Levin, 1993; Hale and Keyser, 2002), and upon noticing that a particular verb occurs with a particular syntactic feature, the learner "reverses" those projection rules to get from that syntactic context to that word's corresponding semantic components (Kako, 1997; Lidz et al., 2004; White, 2015).

Continuing the examples above, a verb's taking a tensed subordinate clause correlates (in English) with that verb being a belief verb (5a), like *think*, *believe*, or *know* (Bolinger, 1968; Stalnaker, 1984; Farkas, 1985; Heim, 1992; Villalta, 2000, 2008; Anand and Hacquard, 2013, among others). This is corroborated by the fact that desire verbs, like *want*, *prefer*, and *order*, which arguably do not have a belief component (though see Heim 1992), and do not take tensed subordinate clauses (5b).

(5) a. John {thinks, believes, knows} that Mary is happy.
b. *John {wants, needs, orders} that Mary is happy.

Assuming this correlation to be cross-linguistically robust, a syntactic bootstrapping account might then posit that learners have some innate projection rule (6a) that they can reverse to get from the fact that *think*, *believe*, and *know* occur with finite complements (6b) to the fact that *think*, *believe*, and *know* have a meaning that involves belief (6c) (De Villiers and De Villiers, 2000; De Villiers and Pyers, 2002; de Villiers, 2005).

(6) **Top-down approach**
a. *Knowledge:* BELIEF $\longrightarrow$ S[+TENSE]
b. *Data:* {think, believe, know} S[+TENSE]
c. *Inference:* BELIEF $\longleftarrow$ {think, believe, know}

The top-down approach makes strong predictions about learners' inferences: labeling is an automatic consequence of noticing a distributional fact—in this case, that some verb takes a tensed subordinate clause.

One difficulty that arises with the top-down approach is that it is not robust to cross-linguistic variation. For instance, suppose that the projection rule in (6a) were innate. One would expect either (i) that all languages show a correlation between a verb's having a belief component and its taking tensed clauses; or (ii) that, if a language does allow non-belief verbs—e.g., desire verbs like *want*—to take tensed clauses, learners might go through a stage where they incorrectly believe that those verbs actually have a BELIEF component. Neither of these possibilities are realized: (i) there are languages, such as German (7) and Spanish (8), where both belief and desire verbs take tensed subordinate clauses; and (ii) in these languages—or at the very least, in German—children do not mistake one type of verb for the other (Perner et al., 2003).

(7) a. Ich glaube, dass Peter nach Hause geht.
I think that Peter to home goes.
b. Ich will, dass Peter nach Hause geht.
I want that Peter to home goes.

(8) a. Creo que Peter va a la casa.
think.1S.PRES that Peter go.PRES.IND to the house.
b. Quiero que Peter vaya a la casa.
want.1S.PRES that Peter go.PRES.SBJ to the house.

The bottom-up approach remedies this issue at the cost of making weaker distributional and developmental predictions. In the bottom-up approach (Alishahi and Stevenson, 2008; Barak et al., 2012, 2013, 2014a,b), learners cluster verbs (9c-i) based on syntactic context (9b), but the clustering mechanism itself does not provide the labels for these clusters. Rather, the learner must notice some correlation between the unlabeled clusters and the sorts of conceptualizations that are triggered by external stimuli when that cluster is instantiated (9c-ii).[1] Then,

[1]The bottom-up approach is similar in form to semantic bootstrapping, in which learners are presumed to have access to the semantics relevant to a particular learning instance (Grimshaw, 1981, 1994; Pinker, 1984, 1989, 1994):

given the cluster each verb falls into (9c-i) and the labeling of that cluster (9c-ii), the learner can make the inference that those verbs have that label (9c-iv).[2]

(9) **Bottom-up approach**

a. *Knowledge:* $\emptyset$

b. *Data:* (BELIEF, {think, believe, know} S[+TENSE])

c. *Inferences*

(i) CLUSTER 1 $\longleftarrow$ {think, believe, know}

(ii) CLUSTER 1 $\longleftrightarrow$ BELIEF

(iii) CLUSTER 1 $\longleftrightarrow$ S[+TENSE]

(iv) BELIEF $\longleftarrow$ {think, believe, know}

(v) BELIEF $\longrightarrow$ S[+TENSE]

The bottom-up approach is thus robust to cross-linguistic variability, since a learning mechanism that implements it can learn arbitrary projection rules—e.g. from BELIEF to tense (9c-v) in English—by noticing a correlation between the cluster and the syntax (9c-iii).[3]

This robustness is also the source of its major problem. To make the inferential step in (9c-ii), the learner must have access to the pairing of a conceptualization—e.g. BELIEF—with a word—e.g. *think*. But there is mounting evidence that, even in contexts that are constructed so as to heavily bias toward activating abstract concepts like BELIEF and DESIRE, propositional attitude meanings are not considered as candidates the majority of the time (Papafragou et al., 2007); and in more naturalistic contexts, they are almost never considered (Gillette et al., 1999; Snedeker and Gleitman, 2004; Gleitman et al., 2005).

If we take this problem of observability seriously, we need a way to resolve the labeling problem, and a bottom-up approach just can't work. For this reason, we pursue a solution that modifies the top-down approach so as to be robust to cross-linguistic variability. We present the outlines of our proposal in Section 3. We implement this proposal in Section 4, and we conduct an experiment using this implementation in Section 5.

3 Our proposal

To reiterate, a major challenge for the standard top-down approach is that the particular syntactic features associated with belief vs. desire verbs differ cross-linguistically. In English, this

in this case, the fact that the conceptual content BELIEF "cooccurs" with the linguistic content {think, believe, know} S[+TENSE]. It differs, however, in the sense that semantic bootstrapping is a theory of how children come to learn the syntax of their language, whereas the bottom-up approach assumes access to the syntax as a prerequisite. Further, the traditional version of the semantic bootstrapping is more like the top-down approach in assuming learners have innate projection rules (though see Connor et al. 2013).

[2]There is of course knowledge that learners are required to have for either approach to work that we are not listing here. For example, both the top-down and bottom-up approaches require (i) that the relevant syntactic structures can be parsed by learners at the relevant developmental stage, and (ii) that the relevant conceptual material is accessible to them at that stage. This second requirement may or may not be met at certain points in development. See Onishi and Baillargeon 2005; Baillargeon et al. 2010 for evidence that this conceptual material is accessible from a very young age.

[3]This is similar to the explanation for how the label itself is associated with the cluster. Indeed, some bottom-up models, such as Alishahi and Stevenson's (2008), explicitly treat the association between the cluster and the concept as of the same type as the association between the cluster and the syntactic feature—i.e. the projection rule (see also Barak et al., 2012, 2013, 2014a,b). This is because they treat both the concept and the syntax as observed features of the verb, which can both be used in forming the cluster in the first place.

In fact, for the purposes of learning a word's meaning, the syntactic features are to some extent superfluous for models that employ the bottom-up approach, since the semantics themselves are observed and can thus contribute to forming a cluster with a particular label. In this sense, the bottom-up approach is essentially a cross-situational word learning model (Yu and Smith, 2007; Smith and Yu, 2008; Yu and Smith, 2012; Medina et al., 2011; Trueswell et al., 2013) with additional context features (cf. Frank et al., 2009).

meaning distinction is tracked by whether the complement clause is tensed or not. But as we saw, in German and Spanish, both belief and desire verbs take tensed complements.

Interestingly, the belief v. desire distinction is still tracked by the syntax of subordinate clauses, albeit via different means. In Spanish (and other Romance languages), it is tracked by the mood of the subordinate clause: belief verbs tend to take subordinate clauses with indicative mood and desire verbs tend to take subordinate clauses with subjunctive mood, exemplified in (8) from Section 2. And in German, the distinction is tracked by whether the complement has verb second (V2) syntax (Truckenbrodt, 2006; Scheffler, 2009): belief verbs tend to allow subordinate clauses with V2 syntax and desire verbs tend not to, exemplified in (10).

(10) a. Ich glaube, Peter geht nach Hause.
  I think Peter goes to home.
 b. *Ich will, Peter geht nach Hause.
  I want Peter goes to home.

We argue that these syntactic features converge at an abstract level and that this convergence could help resolve the cross-linguistic challenge for a top-down approach (cf. Hacquard, 2014; Hacquard and Lidz, submitted).

In particular, belief verbs take subordinate clauses with syntactic hallmarks of declarative main clauses in their respective languages. For instance, in English, belief verbs' subordinate clauses tend to be tensed, just like declarative main clauses in English; in Romance, belief verbs' subordinate clauses tend to have the indicative mood, just like declarative main clauses in Romance; and in German, belief verbs' subordinate clause can have verb second word order, just like German declarative main clauses. Analogously, desire verbs tend to take subordinate clauses that show hallmarks of imperative main clauses.[4]

We propose to exploit this cross-linguistic convergence by invoking what we term *abstract projection rules*. An abstract projection rule is a generalization of the traditional notion of a projection rule discussed in Section 2. Instead of a particular semantic component—e.g. BELIEF—mapping onto a particular syntactic feature value—e.g. S[+TENSE]—in an abstract projection rule, a particular semantic component maps onto a set of unvalued syntactic features. Over the course of learning, learners must learn a valuation of the syntactic features that appear in this abstract projection rule before that rule can be used in syntactic bootstrapping.

The *featural anchor* for an abstract projection rule is a class of syntactic structures that (i) determine how the syntactic features in the abstract projection rule are valued and (ii) are identifiable prior to verb learning. That is, the featural anchor is, in essence, a valuation of the syntactic features listed in the abstract projection rule that is easy to identify.

Based on the correlation mentioned above, we suggest that the featural anchor for the BELIEF projection rule is the declarative main clause and the featural anchor for the DESIRE projection rule is the imperative main clause—i.e. BELIEF tends to project onto whatever syntactic features are instantiated by a language's declarative main clauses and DESIRE tends to project onto whatever syntactic features are instantiated by a language's imperative main clauses. For instance, instead of the projection rule (11), which we saw fails for languages like German and Spanish, a learner might instead have rules of the form in (12).

(11) a. BELIEF $\longrightarrow$ S[+TENSE]
 b. DESIRE $\longrightarrow$ S[-TENSE]

(12) a. BELIEF $\longrightarrow$ DECLARATIVE MAIN CLAUSE
 b. DESIRE $\longrightarrow$ IMPERATIVE MAIN CLAUSE

[4]This latter claim is somewhat more tentative. Interestingly, however, we find that assuming that it is true significantly improves the performance of the model we describe below—at least on English.

Learners must then find the valuations for DECLARATIVE MAIN CLAUSE and IMPERATIVE MAIN CLAUSE specific to their language, at which point they can use (12) as they would projection rules like those in (11).

But why should main clause syntax matter? Is there a principled link between this abstract syntax and the attitude verbs' underlying semantics? We have argued that there is and that the connection comes from the association of particular clause types with different speech acts (see Hacquard, 2014; Hacquard and Lidz, submitted).

Cross-linguistically, languages devote particular clause types to different speech acts: declaratives are typically associated with assertions, imperatives with commands, and interrogatives with questions (Sadock and Zwicky, 1985). Attitude reports are often used to perform indirect speech acts (Searle, 1975), and different attitudes easily lend themselves to different indirect speech acts because of the meaning they express.

For instance, an assertion is an expression of a judgment of truth: if one asserts (13), one commits oneself to the truth of *it is raining*.

(13) It is raining.

And because belief verbs report judgments of truth (Bolinger, 1968; Stalnaker, 1984; Farkas, 1985; Heim, 1992; Villalta, 2000, 2008; Anand and Hacquard, 2013, a.o.), they easily lend themselves to indirect assertions (Urmson, 1952; Hooper, 1975; Simons, 2007; Lewis, 2013; Anand and Hacquard, 2013). A speaker can indirectly assert the content of the complement clause by implicitly endorsing the reported judgment of truth.

(14) **A:** Why are you putting on a rain jacket?
B: John {thinks, said} that it's raining.

Analogously, a command is an expression of a desire: (15a) expresses my desire that you leave.

(15) (You,) leave!

Desire verbs report desire, and hence easily lend themselves to indirect commands. A speaker can indirectly demand that the state of affairs expressed by the complement clause be brought about by implicitly endorsing the reported desire—compare the pragmatic effects of (15) to those of (16).

(16) I want you to leave

To summarize, attitude verbs seem to split into two main classes: those that express a judgment of truth (belief and speech verbs) and those that express a preference (desire and command verbs). This semantic split is reflected both in the syntax, and in the pragmatics.

In the pragmatics, it is reflected by the type of indirect speech acts that these verbs are routinely used for: indirect assertions for belief verbs, indirect commands for desire verbs. Syntactically, this split seems to be tracked by syntactic features of their complement clauses: belief verbs take complements that resemble declarative main clauses—the syntax typically used for assertions—and desire verbs take complements that resemble imperative main clauses—the syntax typically used for commands.

We suggest that learners may exploit these parallels between speech act and clause type, inferring that a verb that takes a complement with "assertive" syntax expresses a judgment of truth, while a verb that takes a complement with "imperative" syntax expresses a preference. In Section 4, we demonstrate how to implement this idea in a learning model, and we apply this implementation to child-directed speech data in Section 5.

4 Implementing our proposal

In this section, we define a probabilistic model that implements the proposal from Section 3. We do this in two steps; first, we describe a base model that clusters verbs based on their syntactic distributions; second, we show how to augment this model with the abstract projection rule and featural anchor proposed in Section 3.

4.1 Base model

Our base model has two components. The first component describes the relationship between a verb's semantic features and its acceptability in syntactic contexts with particular feature valuations. We refer to this component as the *competence model*. The second component describes the relationship between a verb's acceptability in a particular syntactic context and the syntactic contexts it actually occurs in.

4.1.1 The competence model

We base our competence model on White and Rawlins's (2016) model of semantic selection—itself based on White's (2015, Ch. 3) model of syntactic bootstrapping. The competence model has two components: a representation of a verb's semantic components and a representation of projection rules. Both components are represented as probabilities: s_{vk} is the probability that verb v has semantic component k, and p_{kf} is the probability that a semantic component k projects onto syntactic feature f.

Following White and Rawlins, we define the probability d_{vf} that a verb v is acceptable with a particular syntactic feature f in terms of s_{vk} and p_{kf}.

(17) $d_{vf} \equiv 1 - \prod_k 1 - s_{vk}p_{kf}$ where k ranges over semantic components

The definition in (17) follows from the assumption that s_{vk} and p_{kf} are independent; in this case, $s_{vk}p_{kf}$ is the joint probability that verb v has semantic component k and that semantic component k projects onto syntactic feature f. Thus, $1 - s_{vk}p_{kf}$ is the probability that either verb v does not have semantic component k or semantic component k does not project onto syntactic feature f. Again assuming independence, $\prod_k 1 - s_{vk}p_{kf}$ gives the probability that, for all semantic components k, either verb v does not have semantic component k or semantic component k does not project onto syntactic feature f. Finally, (17) itself gives the probability that there is some semantic component k such that verb v has semantic component k and that semantic component k projects onto syntactic feature f, and thus that a verb v is acceptable with syntactic feature f (see White and Rawlins 2016 for an explicit derivation).

4.1.2 The performance model

To complete the base model, we need some way of linking the competence model to the observed data. As it stands, the model only specifies the probability d_{vf} that a particular verb v is acceptable with a particular syntactic feature f. This is importantly distinct from the probability of actually seeing verb v with syntactic feature f.

There are various ways to model the latter probability. In the current case, we define a probability o_{vk} that verb v instantiates semantic component k on any particular observation of a verb. Then, we define the probability $\hat{d}_{vf}$ of seeing verb v with syntactic feature f on any particular observation of the verb.

(18) $\hat{d}_{vf} \equiv 1 - \prod_k 1 - o_{vk}s_{vk}p_{kf}$

The definition in (18) follows from reasoning analogous to that given for (17) in Section 4.1.1.

We assume that each datapoint i consists of a verb v_i and a syntactic feature combination $\mathbf{x}_i$, which represents a sequence of F binary syntactic features as a bit vector of length F, where $x_{if} = 0$ means that the f^{th} syntactic feature has a - value and $x_{if} = 1$ means that the f^{th} syntactic feature has a + value. We then define the likelihood of the i^{th} feature valuation $\mathbf{x}_i$ given $\mathbf{S}, \mathbf{O}, \mathbf{P}$ and the i^{th} verb v_i as in (19), assuming that each feature valuation x_{if} arises via an independent Bernoulli distribution.

(19) $\mathbb{P}(\mathbf{x}_i \mid v_i, \mathbf{S}, \mathbf{O}, \mathbf{P}) = \prod_f \text{Bernoulli}(x_{if}; \hat{d}_{vf}) = \prod_f \hat{d}_{v_i f}^{x_{if}} (1 - \hat{d}_{v_i f})^{1 - x_{if}}$

Assuming that each $\mathbf{x}_i$ is conditionally independent (given $v_i, \mathbf{S}, \mathbf{O}, \mathbf{P}$) of all the other datapoints, the log-likelihood of the entire dataset is given by (20).

(20) $\mathcal{L}(\mathbf{X} \mid \mathbf{v}, \mathbf{S}, \mathbf{O}, \mathbf{P}) = \sum_i \sum_f x_{if} \log \hat{d}_{v_i f} + (1 - x_{if}) \log(1 - \hat{d}_{v_i f})$

Our objective is to find values for $\mathbf{S}$, $\mathbf{O}$, and $\mathbf{P}$ that maximize (20), relative to some set of constraints. For instance, one (somewhat uninteresting) constraint that we need is that, for all v, k, and f, o_{vk}, s_{vk}, and p_{kf} should be between 0 and 1, since they are probabilities.[5] A more interesting constraint comes from implementing our proposal on top of this base model.

4.2 Implementing abstract projection rules and featural anchors

To implement our proposal, we need some way of representing abstract projection rules and featural anchors for those rules. Recall that we proposed the abstract projection rules in (21), where the right hand side of this rule indicates the class of structures—i.e. the featural anchors—one must observe to fix the syntactic feature valuation for that rule.

(21) a. BELIEF $\longrightarrow$ DECLARATIVE MAIN CLAUSE
b. DESIRE $\longrightarrow$ IMPERATIVE MAIN CLAUSE

To allow our model to fix these features, we pretend that declarative and imperative main clauses are themselves embedded under abstract verbs ASSERT and REQUEST, respectively. So every time our model receives a sentence to process it will also receive either a datapoint like (ASSERT, $\mathbf{x}$) or one like (REQUEST, $\mathbf{x}$), where $\mathbf{x}$ encodes the syntactic features of the main clause of that sentence.[6] For instance, in English, the main clause features observed with ASSERT will tend to be that the clause has a subject and tense but no complementizer, and the main clause features observed with REQUEST will tend to be that the clause has neither a subject nor tense nor a complementizer.

Then, we initialize the model in such a way that ASSERT and REQUEST have only a single semantic component each (and no other semantic components) with probability 1. We stipulate that the semantic component ASSERT has with probability 1 is the BELIEF component—i.e. $s_{\text{ASSERT, BELIEF}} = 1$ and $s_{\text{ASSERT, K}} = 0$ if $k \neq$ BELIEF—and that the semantic component that REQUEST has with probability 1 is the DESIRE component—i.e. $s_{\text{REQUEST, DESIRE}} = 1$ and $s_{\text{REQUEST, K}} = 0$ if $k \neq$ DESIRE.

We then disallow the model from raising the probability of any other semantic component for ASSERT or REQUEST over the course of learning (as it does for observed verbs, such as *think*).

[5] Another (soft) constraint that we need pertains to the fact that, as we have stated the optimization problem, $\mathbf{S}$ and $\mathbf{O}$ are not identifiable (in a statistical sense). For arbitrary v and f, if $s_{vk} = x \neq y = o_{vk}$, there is an equivalent model, with respect to (19), wherein $s_{vk} = y \neq x = o_{vk}$.

To remedy this issue we place an independent sparse Beta(0.5, 0.5) prior on s_{vk} and an independent dense Beta(2, 2) prior on o_{vk} for all v, k. This encourages the model to associate values closer to 0 or 1 with s_{vk} and values closer to 0.5 with o_{vk}. Thus, if the model finds that a value near 0 or 1 is necessary, it favors placing it in $\mathbf{S}$ rather than $\mathbf{O}$. Beyond breaking the symmetry of $\mathbf{S}$ and $\mathbf{O}$ with respect to the objective, this also means that the model is encouraged to make confident guesses about which verbs have which semantic components.

[6] We ignore interrogative main clauses for the purpose of our experiment, but they can be treated in an analogous way.

This ensures that the projection rules $\mathbf{p}_{\text{BELIEF}}$ and $\mathbf{p}_{\text{DESIRE}}$ have a strong pressure to have high probability for syntactic features that are observed in declarative main clauses and imperative main clauses, respectively, since this ensures that $\hat{\mathbf{d}}_{\text{ASSERT}} = \mathbf{p}_{\text{BELIEF}}$ and $\hat{\mathbf{d}}_{\text{REQUEST}} = \mathbf{p}_{\text{DESIRE}}$ regardless of how many sentences have been observed.

Crucially, note that we are *not* hard-coding what the projection rules $\mathbf{p}_{\text{BELIEF}}$ and $\mathbf{p}_{\text{DESIRE}}$ look like. These are randomly initialized and change over the course of learning (see Section 4.3 for details). Rather, we set the model up in such a way that where is a strong pressure to have a $\mathbf{p}_{\text{BELIEF}}$ and a $\mathbf{p}_{\text{DESIRE}}$ that give high probability for whatever syntactic features are observed with ASSERT and REQUEST, which will differ across languages.

Relatedly, there is no *a priori* guarantee that this implementation will work even for English, which is why it is useful to test it on real data. While the sorts of clauses that belief and desire verbs take tend to match the feature valuations of declarative and imperative clauses, respectively, they do not do so perfectly: belief verbs like *think* and *know* can take complementizers, which are not found in main clauses, and desire verbs like *want* and *order* take infinitives with subjects, where imperatives tend not to have subjects are do not contain the infinitival *to*. Our implementation can in principle handle such partial matches between clauses, but it is an empirical question whether it can do so on real data—one which we address in our experiment in Section 5.

4.3 Learning algorithm

Many different kinds of learners can be defined to respect this model's assumptions to varying degrees. Here, we define an incremental learner, which observes pairings of verbs and syntactic features one at a time and makes inferences after each observation. This learner is implemented using a form of stochastic gradient descent with adaptive gradient (Duchi et al., 2011). We do not delve into the specifics of this algorithm, though we do give a high-level description of what it is doing.

The learner begins with randomly initialized matrices $\mathbf{S}$, $\mathbf{O}$, and $\mathbf{P}$ with positive values near 0 (except for $\mathbf{s}_{\text{ASSERT}}$ and $\mathbf{s}_{\text{REQUEST}}$, which are constrained as described in Section 4.2). Upon receiving a particular subset of datapoints corresponding to the verbs found in a particular sentence, the learner calculates how likely that datapoint is given the current model using the Bernoulli likelihood function given in (19). The learner then attempts to change the semantic representation for the verb $\mathbf{s}_v$ and the projection rules $\mathbf{P}$ so that they give a higher likelihood to the data. The adaptive gradient piece of the learner ensures that the changes to the verb's semantic representation are not very extreme if the verb has been seen many times before but are potentially extreme if the verb is very infrequent.

5 Experiment

We now apply the learning algorithm described in Section 4.3 to a dataset of English child-directed speech. We begin by describing the dataset and then present the results. All data and code for this experiment are available at github.com/aaronstevenwhite/MainClauseModel.

5.1 Data

We utilize the subcategorization frame data extracted by (White et al., under review) from the Gleason corpus (Gleason, 1980) in CHILDES (MacWhinney, 2014b,a). Gleason is a useful corpus in our case for a couple reasons. First, it contains transcripts for 24 children in the age range that children are acquiring propositional attitude verbs: 2;1 to 5;2 (De Villiers and De Villiers, 2000; De Villiers and Pyers, 2002; de Villiers, 2005). Second, it contains three types of transcripts that are at least somewhat representative of common situations children find themselves in on a daily basis: play contexts, one with the mother and one with the father, and meal contexts, with both the mother and the father.

For every occurrence of a verb in Gleason, White et al. used the MOR and POST morphological analyses (Parisse, 2000) and the MEGRASP dependency parses (Sagae et al., 2007) that ship with some CHILDES corpora to extract the syntactic features listed in (22).

(22)
a. [+/- DIRECT OBJECT]
b. [+/- PREPOSITIONAL PHRASE]
c. [+/- EMBEDDED CLAUSE]
d. [+/- EMBEDDED SUBJECT]
e. [+/- COMPLEMENTIZER]
f. [+/- TENSE]
g. [+/- INFINITIVAL]

This means that each observation is constituted by a verb paired with a string of seven boolean values—one for each valuation of the feature.

For instance, (23) is an example of a sentence found in the dinner transcript for Bobby from the Gleason corpus. This sentence has the valuation in (23a), which would be fed to the model as (23b).[7]

(23) Do you want me to teach you too?
a. [+ DO, - PP, + EMB CLAUSE, - EMB SUBJ, - COMP, - EMB TENSE, + EMB INFINITIVAL]
b. [1, 0, 1, 0, 0, 0, 1]

Only 22 of the 24 children (11 females) have transcripts for both the dinner session and the play session, and so following previous analyses of these data (cf. Ely et al., 2001), we use only data from these children. For each of these 22 children, we combine the data extracted from the dinner and play contexts. Across children, the mean number of sentences in each dataset is 632.7 (median: 625, IQR: [544, 729]), and the mean number of verbs in each dataset is 1601.2 (median: 1555.5, IQR: [1358.5, 1824.75]).

5.2 Fitting

We apply the algorithm described in Section 4.3 to each of the 22 datasets, randomly selecting a sentence to reveal at each time step until the model has seen 20,000 sentences total. Given that the average number of sentences in the transcript is 632.7 and assuming this number is a reasonable lower bound on the number of sentences a child hears in a day, this simulates approximately a month's worth of input (as an upper bound).

We repeat this procedure 10 times for each dataset. All reported results are based on averages over these 10 runs. For each run, we set the total number of semantic components to eight, which is the total number of unique syntactic feature valuations found across the 22 datasets. Two of these semantic components are, by necessity, reserved for the BELIEF and DESIRE components, which are the only components we report on here.

5.3 Results

Figure 1 shows the median probability that BELIEF and DESIRE semantic components project onto different syntactic features as a function of the number of total sentences seen. The dark

[7] This extraction method necessarily makes certain decisions about the syntactic structure of the sentence—e.g. that *me* in (23) is a direct object and not an embedded subject. (Other reasonable annotation are that *me* is both a direct object and an embedded subject or that it is only an embedded subject.) Such decisions are unavoidable and where there is one to make White et al. follow the dependency labels available from parse itself as closely as possible.

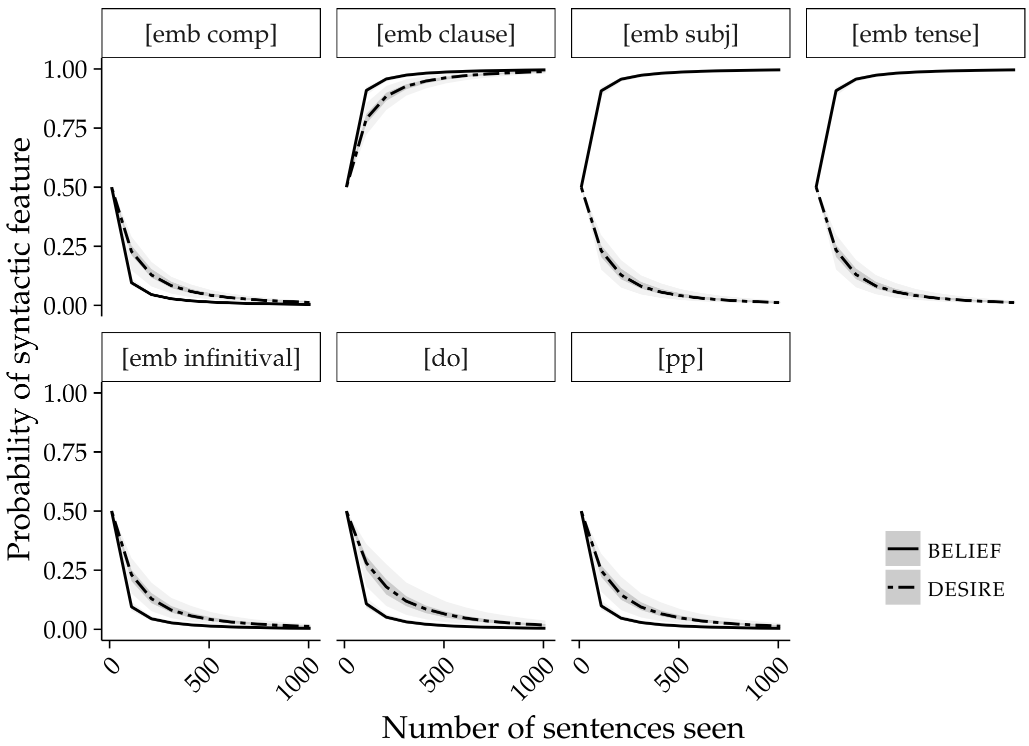

Figure 1: Median probability that BELIEF and DESIRE semantic components project onto different syntactic features as a function of the number of total sentences seen.

shading shows the interquartile range over the 22 datasets, and the light shading show the minimum and maximum over the 22 datasets.

We see that the model robustly learns that, in English, BELIEF projects onto tensed subordinate clauses with a subject but no complementizer and that DESIRE projects onto untensed subordinate clauses without a subject or complementizer. Further, this learning happens extremely quickly: by 1000 sentences—roughly, two days worth of input—the model has converged to the aforementioned feature probabilities. This is almost certainly a product of the fact that main clauses are necessarily extremely common.

Figure 2 shows the median probability of BELIEF and DESIRE semantic components for the ten most frequent propositional attitude verbs (in order of frequency in Gleason) as a function of the number of total sentences seen (whether or not those sentences contained the verb in question or not). The dark shading shows the interquartile range over the 22 datasets, and the light shading show the minimum and maximum over the 22 datasets. (Note that the scale on the x-axis of Figure 2 is an order of magnitude larger than that of Figure 1.)

We see that the model robustly learns that *think* has a BELIEF component and that *want* has a DESIRE component. It also robustly learns that *want* does not have a BELIEF component and, except in the most extreme case, that *think* does not have a DESIRE component (and even then, it assigns a 50% probability to *think* having a DESIRE component).

Further, the algorithm converges to the solutions for *think* and *want* relatively quickly. We see that, by about 7,500 sentences, the algorithm has learned that *want* has a DESIRE component and that, by about 10,000 sentence, it has learned that *want* does not have a BELIEF component. The time to convergence is similar for *think*, though it takes slightly longer. This difference is likely a function of the fact that *think* is approximately 25% less frequent than *want*.

Turning now to the other verbs in Figure 2, we see that our algorithm does well in labeling *know* and *tell* with a BELIEF component on a majority of the datasets. It shows much higher variability across datasets with *remember* and *see*, and its performance with *say* is poor except

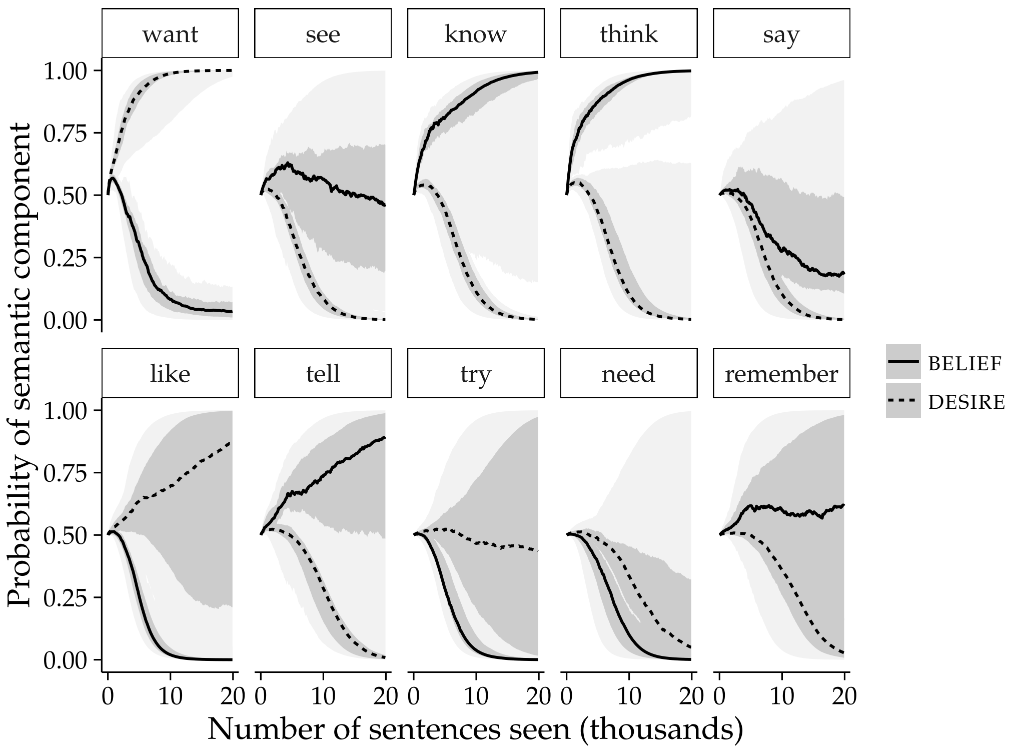

Figure 2: Median probability of BELIEF and DESIRE semantic components for ten most frequent attitude verbs as a function of the number of total sentences seen.

in some extreme cases.

A similar variability is found for verbs like *like* and *need*, which one might expect to be labeled with a desire meaning. On a (slim) majority of the transcripts, the model assigns high probability to *like* having a DESIRE component, but it does not do this for a significant proportion. And though *need* tends to be assigned higher probability for DESIRE than for BELIEF, for most transcripts this probability is low. (It is somewhat unclear whether *try* should get a desire meaning or not, but insofar as it should, it is something of an intermediate case between *like* and *need*.)

What appears to be driving this variability is that some of the 22 children's transcripts contain very few occurrences of some verbs with embedded clauses. For instance, while *know*, *tell*, and *like* occur with embedded clauses on average 24%, 27%, 37% of the time, repectively, *see* and *say* occur with embedded clauses on average 15% and 14% of the time, repectively.[8] This is corroborated when looking at how the proportion of embedded clauses found in a transcript affects the final state of the model.

Figure 3 plots the mean probability of the BELIEF and DESIRE semantic components after seeing final sentence for ten most frequent propositional attitude verbs (in order of frequency in Gleason) as a function of the proportion of times that verb is found with an embedded clause in a particular transcript. (The y-axis is logit-scaled.) Here, we see that the relative frequency of embedded clauses for a particular a verb in a transcript is strongly related to the probability that the model assigns to the verb having either a BELIEF component or a DESIRE component (Spearman's ρ=0.84).

[8] *Remember* and *need* break this trend to some extent, since they occur with embedded clauses on average 30% and 26% of the time, respectively. Upon inspection of the dataset and the dependency parses on which it is based, this appears to be driven by a combination of poor parses for a signficant portion of sentences containing *remember* and *need*.

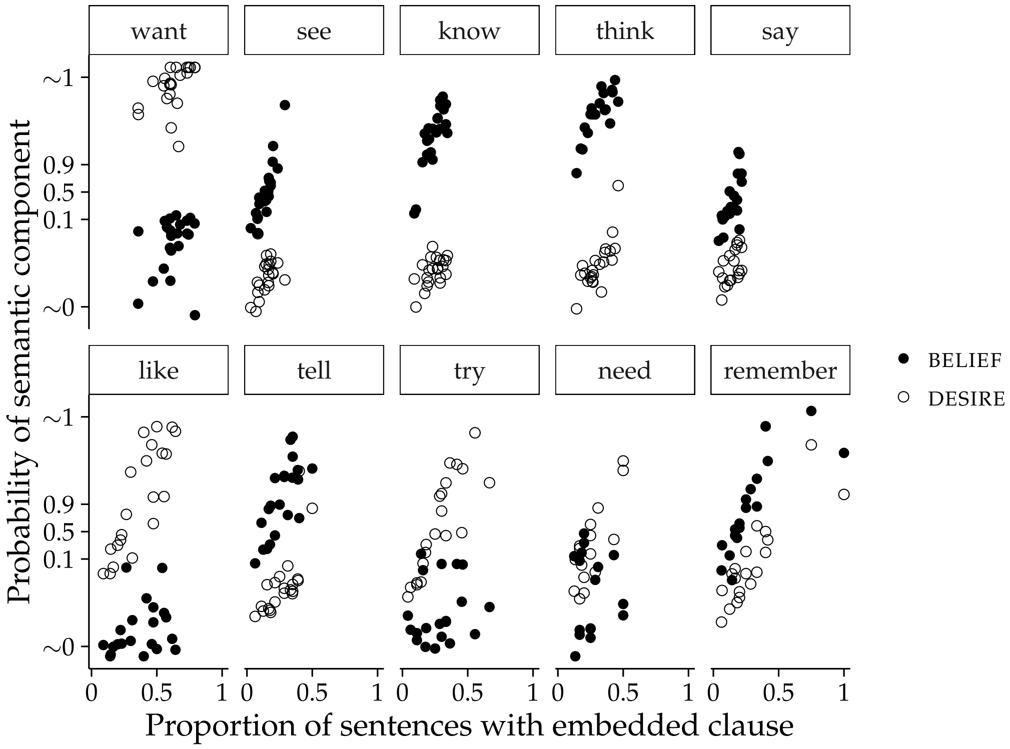

Figure 3: Mean probability of BELIEF and DESIRE semantic components after seeing final sentence for ten most frequent attitude verbs as a function of the proportion of times that verb is found with an embedded clause in a particular transcript.

5.4 Discussion

We have shown that our model works well for labeling core cases of belief and desire verbs, such as *think*, *know*, *want*, and *tell*, but that it shows variability for other verbs, such as *say*, *like*, and *need*. We gave evidence that this variability is related to the relative frequency of clausal embedding for a particular verb.

This raises two questions. First, to what extent are the empirical relative frequencies found in each dataset indicative of each child's experience? Each transcript is only a small sample of children's experience, and so the relative frequencies plotted in Figure 3 may well not accurately reflect this experience. This matters, especially in cases where the average relative frequency of subordinate clauses for a particular verb is near 0—e.g. for *say*—since it may well be that children's actual experience includes many more instances of the verb with a subordinate clause.[9]

Second, to what extent is this pattern really a fact about relative frequency of a particular syntactic feature paired with a verb and not overall frequency of the verb? For instance, even if a verb shows up only rarely with a subordinate clause, does this necessarily mean that the child will not learn that the verb has, e.g., a belief or desire component? One possibility is that it just takes a substantially longer time than we have simulated here to learn that such verbs have a belief or desire component.

[9]A potential confound in these transcripts—particularly for *say*—is that the play sessions include book reading, which involves many cases of quotation—e.g. *say 'hi'*. Depending on the amount of book reading input a particular child receives, this may warp the distribution of communicative verbs like *say*.

6 General discussion

The main substantive addition we make to the theory of syntactic bootstrapping in this paper is the notion of a featural anchor, which is itself a class of structures known prior to the selection of a projection rule. One question this addition raises is to what extent learners can easily discover the featural anchor itself. A deflationary response to our proposal might argue that we have merely pushed the job of verb learning back to discovering which classes of structures constitute featural anchors. Isn't the job of figuring out which syntactic features are indicative of a particular anchor just as hard as learning a verb itself?

Yes and no. Yes, because it is true that, at the end of the day, one must identify some structure as a declarative or imperative main clause, and this must presumably be done by perceiving that, in using a particular structure, an utterer of a declarative main clause intends the utterance to be taken as part of a particular conversational move—such as an assertion (Austin, 1975; Stalnaker, 1978). That is, the learner must be able to identify the illocutionary force intended for the utterance.

No, because illocutionary force is a concept that is presumably prerequisite to learning a language in a first place. Indeed, children appear to be adept at recognizing an utterance's illocutionary force quite early (Spekman and Roth, 1985). This is to say that, though language is clearly not *for* communication, the data a learner uses to learn their language tends to come wrapped in communicative acts, which the learner presumably has no problem perceiving as such.

And no, because the mapping between illocutionary force and syntactic structure is relatively stable within a language: assertions, at least as conveyed by clauses, are conveyed by clauses with the same syntactic features—in English, [+TENSE, -COMP, ...]. And insofar as an assertion is not conveyed by a clause—such as when it is conveyed by a polarity particle or a fragment—we submit that, if a learner has enough syntactic knowledge to represent a clause as a set of features, they have enough syntactic knowledge to represent that the valuation of those features is dependent on the fact that that clause is a syntactic object in the first place.

This question of recognizing a syntactic class of complements, such as a clause, is related to the issue we saw our algorithm having with nominal and propositional anaphor complements, like *so*. These complements cannot be valued for the same sorts of features that a clause can be—indeed, they don't appear to be valued as such at any level of syntactic representation (Hankamer and Sag, 1976)—but as it stands our model views them as simply unvalued. But rather than view them as unvalued for these features, it seems that the distinction in syntactic class must be baked into the model itself. That is, the syntactic features the model pays attention to in making an inference from a particular piece of data must be dependent on the syntactic class involved in that datum. In the abstract, we need to incorporate some decision tree-like representation into the model. One way this might be implemented is by employing a likelihood function that incorporates a hurdle model. (See White et al. to appear for a recent use of hurdle models in a related domain.)

Beyond providing a case for a decision tree-like structure for syntactic feature valuation, the case of propositional anaphors may also suggest that the mechanisms used by syntactic bootstrapping to infer a verb's meaning may need to incorporate a notion of semantic type over and above that given by syntactic type. (See White and Rawlins 2016 for evidence that semantic type signatures can be extracted from syntactic distribution.)

Another question that arises is to what extent there are other abstract projection rules and featural anchors. There seems to us to be at least one further candidate: factive and interrogative verbs like *know* and *wonder*.

Dudley et al. (in prep) show that a large part of children's experience with *know* is in the frame *do you know Q?* We would like to suggest that, here again, the child may be able exploit the syntactic parallels between direct and indirect speech acts: *know* and *wonder* are used to ask indirect questions.

(24) a. Do you know where the keys are?
b. I was wondering where you put my keys.

Our suggestion is that, analogous to what we claim for belief and desire verbs, children might infer from this parallel that the meaning of *know* must be one that relates the subject to the answer of that question. In future work, we aim to investigate this possibility.

7 Conclusion

In this paper, we proposed a novel solution to the labeling problem in syntactic bootstrapping that augments the standard top-down approach to syntactic bootstrapping with the concepts of an *abstract projection rule* and a concomitant *featural anchor*. We motivated this proposal by noting that neither the top-down nor the bottom-up approaches solve the labeling problem for belief and desire verbs: the top-down approach is brittle in the face of cross-linguistic variability, while the bottom-up approach makes unrealistic assumptions about the data learners have access to.

We showed that our proposed solution can deal with the labeling problem given theoretically justified featural anchors for particular labels, and using belief and desire predicates as a case study, we implemented a computational model that incorporates the labeling mechanism we propose. We presented a proof-of-concept fit of this model to data derived from child-directed speech and showed that our model works well for labeling core cases of belief and desire verbs, such as *think*, *know*, *want*, and *tell*.

References

Alishahi, Afra, and Suzanne Stevenson. 2008. A computational model of early argument structure acquisition. *Cognitive Science* 32:789–834.

Anand, Pranav, and Valentine Hacquard. 2013. Epistemics and attitudes. *Semantics and Pragmatics* 6:1–59.

Austin, John Langshaw. 1975. *How to do things with words*. Oxford university press.

Baillargeon, Renée, Rose M. Scott, and Zijing He. 2010. False-belief understanding in infants. *Trends in cognitive sciences* 14:110–118.

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

Bolinger, Dwight. 1968. Postposed main phrases: an English rule for the Romance subjunctive. *Canadian Journal of Linguistics* 14:3–30.

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

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

Connor, Michael, Cynthia Fisher, and Dan Roth. 2013. Starting from scratch in semantic role labeling: Early indirect supervision. In *Cognitive aspects of computational language acquisition*, 257–296. Springer.

De Villiers, Jill G., and Peter A. De Villiers. 2000. Linguistic determinism and the understanding of false belief. *Children's Reasoning and the Mind* 191–228.

De Villiers, Jill G., and Jennie E. Pyers. 2002. Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development* 17:1037–1060.

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

Dudley, Rachel, Meredith Rowe, Valentine Hacquard, and Jeffrey Lidz. in prep. On the syntactic distribution of mental state verbs in child-directed speech.

Ely, Richard, Jean Berko Gleason, Ann MacGibbon, and Elena Zaretsky. 2001. Attention to language: Lessons learned at the dinner table. *Social Development* 10:355–373.

Farkas, Donka. 1985. *Intensional descriptions and the Romance subjunctive mood*. Taylor & Francis.

Frank, Michael C., Noah D. Goodman, and Joshua B. Tenenbaum. 2009. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* 20:578–585.

Gillette, Jane, Henry Gleitman, Lila Gleitman, and Anne Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73:135–176.

Gleason, Jean Berko. 1980. The acquisition of social speech routines and politeness formulas. In *Language: Social Psychological Perspectives*, ed. H. Giles, W.P. Robinson, and P.N. Smith, 21–27. Oxford: Pergamon Press.

Gleitman, Lila. 1990. The structural sources of verb meanings. *Language acquisition* 1:3–55.

Gleitman, Lila R., Kimberly Cassidy, Rebecca Nappa, Anna Papafragou, and John C. Trueswell. 2005. Hard words. *Language Learning and Development* 1:23–64.

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.

Grimshaw, Jane. 1994. Lexical reconciliation. *Lingua* 92:411–430.

Gruber, Jeffrey Steven. 1965. Studies in lexical relations. Doctoral Dissertation, Massachusetts Institute of Technology.

Hacquard, Valentine. 2014. Bootstrapping attitudes. In *Semantics and Linguistic Theory*, volume 24, 330–352.

Hacquard, Valentine, and Jeffrey Lidz. submitted. Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics .

Hale, Ken, and Samuel Jay Keyser. 2002. *Prolegomena to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 391–428.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of semantics* 9:183–221.

Hooper, Joan B. 1975. On Assertive Predicates. In *Syntax and Semantics*, ed. John P. Kimball, volume 4, 91–124. New York: Academy Press.

Kako, Edward. 1997. Subcategorization Semantics and the Naturalness of Verb-Frame Pairings. *University of Pennsylvania Working Papers in Linguistics* 4:11.

Landau, Barbara, and Lila R. Gleitman. 1985. *Language and experience: Evidence from the blind child*, volume 8. Harvard University Press.

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

Lewis, Shevaun. 2013. Pragmatic enrichment in language processing and development. Doctoral Dissertation, University of Maryland.

Lidz, Jeffrey, Henry Gleitman, and Lila Gleitman. 2004. Kidz in the 'hood: Syntactic bootstrapping and the mental lexicon. In *Weaving a Lexicon*, ed. D.G. Hall and S.R. Waxman, 603–636. Cambridge, MA: MIT Press.

MacWhinney, Brian. 2014a. *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

MacWhinney, Brian. 2014b. *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Medina, Tamara Nicol, Jesse Snedeker, John C. Trueswell, and Lila R. Gleitman. 2011. How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences* 108:9014–9019.

Onishi, Kristine H., and Renée Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308:255–258.

Papafragou, Anna, Kimberly Cassidy, and Lila Gleitman. 2007. When we think about thinking: The acquisition of belief verbs. *Cognition* 105:125–165.

Parisse, Christophe. 2000. Automatic disambiguation of morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments, & Computers* 32:468–481.

Perner, Josef, Manuel Sprung, Petra Zauner, and Hubert Haider. 2003. Want That is Understood Well before Say that, Think That, and False Belief: A Test of de Villiers's Linguistic Determinism on German–Speaking Children. *Child development* 74:179–188.

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.

Pinker, Steven. 1994. How could a child use verb syntax to learn verb semantics? *Lingua* 92:377–410.

Sadock, Jerrold M., and Arnold M. Zwicky. 1985. Speech act distinctions in syntax. *Language Typology and Syntactic Description* 1:155–196.

Sagae, Kenji, Eric Davis, Alon Lavie, Brian MacWhinney, and Shuly Wintner. 2007. High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, 25–32. Association for Computational Linguistics.

Scheffler, Tatjana. 2009. Evidentiality and German attitude verbs. *University of Pennsylvania Working Papers in Linguistics* 15.

Searle, John. 1975. Indirect speech acts. In *Speech Acts*, ed. P. Cole and J.L. Morgan, volume 3 of *Syntax and Semantics*, 59–82. New York: Academic Press.

Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117:1034–1056.

Smith, Linda, and Chen Yu. 2008. Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106:1558–1568.

Snedeker, Jesse, and Lila Gleitman. 2004. Why it is hard to label our concepts. In *Weaving a Lexicon*, 257–294. Cambridge, MA: MIT Press.

Spekman, Nancy J., and Froma P. Roth. 1985. Preschool children's comprehension and production of directive forms. *Journal of Psycholinguistic Research* 14:331–349.

Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics (New York Academic Press)* 9:315–332.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge University Press.

Truckenbrodt, Hubert. 2006. On the semantic motivation of syntactic verb movement to C in German. *Theoretical Linguistics* 32:257–306.

Trueswell, John C., Tamara Nicol Medina, Alon Hafri, and Lila R. Gleitman. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology* 66:126–156.

Urmson, James O. 1952. Parenthetical verbs. *Mind* 61:480–496.

Villalta, Elisabeth. 2000. Spanish subjunctive clauses require ordered alternatives. In *Semantics and Linguistic Theory*, volume 10, 239–256.

Villalta, Elisabeth. 2008. Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy* 31:467–522.

de Villiers, Jill G. 2005. Can Language Acquisition Give Children a Point of View? In *Why language matters for theory of mind*, ed. Janet W. Astington and Jodie A. Baird, 186–219.

White, Aaron Steven. 2015. Information and incrementality in syntactic bootstrapping. Doctoral Dissertation, University of Maryland.

White, Aaron Steven, Valentine Hacquard, and Jeffrey Lidz. accepted. Semantic information and the syntax of propositional attitude verbs. *Cognitive Science* .

White, Aaron Steven, Valentine Hacquard, Philip Resnik, and Jeffrey Lidz. under review. The contextual modulation of semantic information .

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

White, Aaron Steven, Kyle Rawlins, and Benjamin Van Durme. to appear. The semantic proto-role linking model. *Proceedings of 15th Conference of the European Chapter of the Association for Computational Linguistics* .

Yu, Chen, and Linda B. Smith. 2007. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18:414–420.

Yu, Chen, and Linda B. Smith. 2012. Modeling cross-situational word–referent learning: Prior questions. *Psychological review* 119:21.