

Semantic information and the syntax of propositional attitude verbs*

Aaron Steven White
Johns Hopkins University

Valentine Hacquard
University of Maryland

Jeffrey Lidz
University of Maryland

Abstract

This paper explores the granularity with which a word’s semantic properties are recoverable from its syntactic distribution, taking propositional attitude verbs (PAVs), such as *think* and *want*, as a case study. Three behavioral experiments aimed at quantifying the relationship between PAV semantic properties and PAV syntactic distribution are reported. Experiment 1 gathers a measure of PAV syntactic distributions using an acceptability judgment task. Experiments 2 and 3 gather measures of semantic similarity between those same PAVs using a generalized semantic discrimination (triad or “odd man out”) task and an ordinal (likert) scale task, respectively. Two kinds of analyses are conducted on the data from these experiments. The first compares both the acceptability judgments and the semantic similarity judgments to classifications of PAVs derived from traditional distributional analysis. The second kind compares the acceptability judgments to the semantic similarity judgments directly. Through these comparisons, we show that there is quite fine-grained information about PAV semantics in PAV syntactic distributions—whether one considers the sorts of discrete qualitative classifications that linguists traditionally work with or the sorts of continuous quantitative classifications that can be derived experimentally.

1 Introduction

Theoretical linguists have long been interested in propositional attitude verbs—e.g., *want*, *think*, and *know*—for both their syntactic properties and their semantic properties. These verbs are syntactically interesting because, as a class, they take a wide variety of clausal complements. For instance, *think* and *know* take tensed clausal complements, and *want* takes untensed clausal complements.

- (1) Mary {thought, knew} that John was happy.
- (2) Mary wanted John to be happy.

They are semantically interesting because even superficially quite similar verbs, such as *think* and *know*, can have strikingly different semantic properties—e.g., distinct patterns of entailment. Neither (3a) nor (3b) imply either (5a) or (5b), but both (4a) and (4b) (only) imply (5a).

*We are grateful to audiences at the University of Maryland, Johns Hopkins University, UC Santa Cruz, and NELS 43 as well as Philip Resnik, Naomi Feldman, Hal Daumé III, Alexander Williams, Norbert Hornstein, Colin Phillips, Erin Eaker, Shevaun Lewis, Kaitlyn Harrigan, Naho Orita, Rachel Dudley, Morgan Moyer, Tom Grano, Pranav Anand, Jane Grimshaw, Kyle Rawlins, Ben Van Durme, Jill de Villiers, John Grinstead, Chris Kennedy, Keir Moulton, Paul Portner, Aynat Rubinstein, Meredith Rowe, Florian Schwarz, Toby Mintz, Florian Jaeger, Noah Goodman, Jesse Snedeker, and Elizabeth Bogal-Allbritten for useful comments on the theory, experiments, and modeling contained in this paper. This work was supported by NSF BCS DDRIG grant 1456013, NSF BCS grant 1124338, and NSF DGE IGERT grant 0801465. This manuscript, whose previous title was “Projecting attitudes” was revised and resubmitted to *Cognitive Science* and is currently under review. Comments are welcome and should be addressed to aswhite@jhu.edu.

- (3) a. Mary thought that John was happy.
b. Mary didn't think that John was happy.
- (4) a. Mary knew that John was happy.
b. Mary didn't know that John was happy.
- (5) a. John was happy.
b. John wasn't happy.

Language acquisitionists have also long been interested in propositional attitude verbs, though from a different angle: among other interesting sociocognitive properties, these verbs, in contrast to action verbs like *run* and *kick*, are not associated with concepts that have perceptual correlates (Landau and Gleitman, 1985; Gleitman, 1990). This *problem of observability* was key in Gleitman's (1990) argument for *syntactic bootstrapping*, wherein a learner uses a word's syntactic context in acquiring its meaning (Brown, 1957, 1973; Macnamara, 1972; Naigles, 1990, 1996; Naigles et al., 1993; Fisher, 1994; Fisher et al., 1994; Waxman and Markow, 1995; Lidz et al., 2004; Waxman and Lidz, 2006; Fisher et al., 2010).

These two traditions have largely remained separate, despite having what we believe to be closely aligned goals.¹ On the one hand, the theoretical literature has focused on understanding the fine-grained relationships that exist between word meaning and syntactic structure, without much thought to whether these relationships are actually robust enough to support learning. On the other hand, the acquisition literature has focused on how only very few syntactic distinctions—generally, the distinction between tensed and untensed clausal complements—are leveraged in syntactic bootstrapping. But if the problem of observability is as dire as Gleitman suggests—a view which is supported by much subsequent literature (Gillette et al. 1999; Snedeker and Gleitman 2004; Papafragou et al. 2007 among others)—understanding the correlational strength between the syntax and fine-grained aspects of meaning is crucial.

Our goal in this paper is to test the limits of syntactic bootstrapping by quantitatively assessing correlations between syntax and word meaning in the domain of propositional attitude verbs. We do this in two parts. In the first, we assess whether the fine-grained semantic properties that are discussed in the theoretical literature are in fact predictable based purely on propositional attitude verb syntactic distributions. And to the extent that they are, we aim to find out which syntactic structures (if any) are predictive of them. To do this, we collect a measure of propositional attitude verbs' syntactic distributions using an acceptability judgment-based methodology developed by Fisher et al. (1991). We show that the classifications laid out in the theoretical literature are quite well predicted by syntactic distributions and that the syntactic structures that are predictive of those properties generally match those suggested in the literature. This part is aimed mainly at the acquisitionists interested in generating hypotheses about what syntactic features learners might use in syntactic bootstrapping.

In the second part, we assess the extent to which the semantic properties discussed in the theoretical literature exhaust the semantic information carried in propositional attitude verb syntactic distributions. To do this, we gather an independent measure of verbs' semantics—verb similarity judgments (cf. Fisher et al. 1991; Schwanenflugel et al. 1994, 1996; Lederer et al. 1995)—and ask whether the theoretical semantic properties statistically mediate the relationship between the syntactic distributions and this measure. The idea is that, insofar as the similarity judgments are predictable from syntactic distributions even after controlling for the theoretical semantic properties, we have evidence of further semantic properties that are tracked by syntactic distributions.

¹We are, of course, not the first to say this about propositional attitude verbs (see Naigles 2000 and references therein). As far as we can tell, however, little has been done to actually link these literatures in any systematic way.

We show that there is indeed substantial evidence for such further syntactically tracked semantic properties. This part is aimed mainly at theoreticians interested in augmenting their methodological toolbox, but we believe it is also useful for acquisitionists trying to understand how much learning could in fact be squeezed out of the syntax.

We begin in Section 2 with an overview of relevant theoretical literature on propositional attitude verbs, pointing out areas of connection with the acquisition literature. In Section 3, we present our acceptability judgment experiment, whose design draws heavily on the prior work discussed in Section 2. We then use the data from this experiment to predict the semantic properties discussed in Section 2. In Section 4, we present two similarity judgment experiments, which we compare against both the acceptability judgment data from Section 3 and the semantic properties from Section 2. In Section 5, we conclude.

2 Propositional attitude verb syntax and semantics

In this section, we present a brief overview of the literature on the syntax and semantics of propositional attitude verbs. Each subsection corresponds to a semantic property we attempt to predict in Section 3. Where possible, we point out cases where the property, or a related property, has been studied in the acquisition literature. Beyond laying out the properties themselves, our aim in this section is to show (i) how these semantic properties might map to syntactic distributions and (ii) that there is sufficient uncertainty in these mappings to warrant the quantitative assessments we carry out in Sections 3 and 4.

2.1 Representationality and preferentiality

Perhaps the most well-known semantic distinction among propositional attitude verbs is that between verbs that express beliefs—or represent “mental pictures” or “judgments of truth” (Bolinger, 1968)—and those that express desires—or more generally, orderings on states of affairs induced by, e.g. commands, laws, preferences, etc. (Bolinger, 1968; Stalnaker, 1984; Farkas, 1985; Heim, 1992; Villalta, 2000, 2008; Anand and Hacquard, 2013, a.o.). Within the first class, which we henceforth refer to as the representational, fall verbs like *think* and *know*; and within the second class, which we henceforth refer to as the preferential, fall verbs like *want* and *order*.

There appear to be various aspects of the syntactic distribution that roughly track this distinction in English. One well-known case is finiteness: representational tend to allow finite subordinate clauses (6a) but not nonfinite ones (6b); preferential tend to allow nonfinite subordinate clauses (7b) but not finite ones (7a).²

- (6) a. Bo thinks that Jo went to the store.
b. *Bo thinks Jo to go to the store.
- (7) a. *Bo wants that Jo went to the store.
b. Bo wants Jo to go to the store.

This correlation is quite well studied in the acquisition literature, with particular focus on how it relates to theory of mind (Wimmer and Perner, 1983; de Villiers, 1995, 2005; De Villiers, 2007;

²This correlation is corroborated to some extent in Barak et al. 2012, 2013, 2014a,b using Alishahi and Stevenson’s (2008) computational model, though caution is required here since they investigate only a small set of verbs and ignore various complications inherent to this distinction (discussed below). See White et al. under review for discussion of the various issues inherent to their model and ways that these issues can be surmounted. See also White 2015 and White and Rawlins to appear for much larger scale investigations that take these complications into account.

De Villiers and De Villiers, 2000; De Villiers and Pyers, 2002; Perner et al., 2003; Lewis, 2013). Children tend to reject sentences like (6a) when they report false beliefs—e.g., if Jo didn't go to the store—but not sentences like (7a) when they report desires that are counter to fact. The relative difficulty with sentences like (6a) has been blamed on conceptual difficulty with false belief (Perner et al., 2003), syntactic difficulty with finite complements (de Villiers, 1995, 2005; De Villiers, 2007; De Villiers and De Villiers, 2000; De Villiers and Pyers, 2002), or pragmatic difficulty tied to the assertivity of belief reports (Lewis, 2013). (See below for more on assertivity.)

There are two important things to note about the representational-preferential distinction. First, though this distinction is often discussed as though it were mutually exclusive, some verbs appear to fall into both categories, and suggestively, show up in both frames. For instance, *hope p* involves both a desire that *p* come about and the belief that *p* is possible (Portner, 1992; Scheffler, 2009; Anand and Hacquard, 2013; Hacquard, 2014, but see also Portner and Rubinstein 2013), and it occurs in both finite (8a) and nonfinite (8b) syntactic contexts.

- (8) a. Bo hopes that Jo went to the store.
 b. Bo hopes to go to the store.

Harrigan (2015) and Harrigan et al. (2016) show that this distribution of clausal complements affects children's interpretation of *hope*. When *hope* occurs with a finite complement, children's interpretations show properties similar to the representational *think*—e.g., they overgeneralize that *hope*, like *think*, is used to report true beliefs—but when it occurs with a nonfinite complement children's interpretations show properties similar to the preferential *want*.

Second, the link between representationality and finiteness is just a tendency. Some verbs plausibly classed as representationals allow nonfinite subordinate clauses (9a)/(9b), and others plausibly classed as preferentials allow subordinate clauses that look finite (9c).³ The roughness of this correlation is perhaps not surprising since not all languages track representationality with tense: for instance, various Romance languages track the distinction with mood—representationals tending to take indicative mood and preferentials tending to take subjunctive mood (Bolinger, 1968; Hooper, 1975; Farkas, 1985; Portner, 1992; Giorgi and Pianesi, 1997; Giannakidou, 1997; Quer, 1998; Villalta, 2000, 2008, a.o.).

- (9) a. Bo believes Jo to be intelligent.
 b. Bo claims to be intelligent.
 c. Bo demanded that Jo go to the store.

But though the correlation between representationality and tense is imperfect, even in English, finiteness does not appear to be the only associated syntactic (distributional) property. Also relevant appears to be a distinction in whether the verb's subordinate clause can be fronted—or in Ross's (1973) terms, S-lifted.⁴ At least some representationals' subordinate clauses (10) appear to be able to undergo S-lifting, but many preferentials' subordinate clauses (11) cannot (Bolinger, 1968).⁵

³Whether (9c) involves a finite subordinate clause is to some extent dependent on whether what is often called the English subjunctive involves tense. On the one hand, the complementizer *that* is the same one that occurs with tensed subordinate clauses, but on the other, the verb shows up in its base (untensed) form.

⁴There is a further distinction in the literature made between S-lifts involving first person and third person propositional attitude verb subjects (Reinhart, 1983; Asher, 2000; Rooryck, 2001). We incorporate this first-third distinction into our experiment, but the data regarding this syntactic distinction are murky at best.

⁵Not all representationals allow S-lifting—e.g., *doubt*. This is likely because the availability of S-lifting for a particular verb is conditioned by other semantic and pragmatic properties it has. See the discussion of *assertivity* below.

- (10) Jo already went to the store, I {think, believe, suppose, hear, see}
- (11) a. *Bo already went to the store, I {want, need, demand}.
- b. *Bo to go to the store, I {want, need, order}.

S-lifting may well be quite important for learning whether a verb is a representational: Diessel and Tomasello (2001) find that many of children’s early uses of representationals like *think* show up in S-lifting structures.

2.2 Factivity

The representationality distinction is cross-cut by another common distinction: factivity (Kiparsky and Kiparsky, 1970; Karttunen, 1971; Horn, 1972; Hooper, 1975). Factivity is defined in terms of its discourse effects. Very roughly, a verb is factive if upon uttering a sentence containing a factive verb with a subordinate clause, a speaker takes the content of the subordinate clause for granted regardless of propositional operators placed around the propositional attitude verb: in particular, negation (13b)/(12b) or questioning (13c)/(12c). For instance, each sentence in (12) commits the speaker to (14) being true, but modulo the context, the sentences in (13) do not. That is, in uttering the sentences in (12), the speaker presupposes (14) (Stalnaker, 1973). This suggests that *know*, *love*, and *hate* are factive, while *think*, *believe*, and *say* are not.

- (12) a. Bo {knew, loved, hated} that Jo went to the store.
- b. Bo didn’t {know, love, hate} that Jo went to the store.
- c. Did Bo {know, love, hate} that Jo went to the store?
- (13) a. Bo {thought, believed, said} that Jo went to the store.
- b. Bo didn’t {think, believe, say} that Jo went to the store.
- c. Did Bo {think, believe, say} that Jo went to the store?
- (14) Jo went to the store.

Factivity truly cross-cuts the representationality distinction in that there are verbs representing all four possible combinations: (i) representational (cognitive) factives, like *know*, *realize*, and *understand*, (ii) preferential (emotive) factives, like *love* and *hate*, (iii) representational nonfactives, like *think* and *say*, and (iv) preferential nonfactives, like *want* and *prefer*.

One question that arises here is whether, given the existence of representational+preferential verbs like *hope*, there could also be such representational+preferential factives. In a certain sense, this may be the case for the emotive factives, since it seems like sentences containing them imply that the holder of the emotion also believes the subordinate clause to be true. If all preferential factives are emotive (and show this behavior), this might suggest that there are no non-representational factives.

One must tread carefully here, however, since not all entailments need be encoded in the meaning of the verb—i.e. this belief entailment could plausibly arise via the same sorts of pragmatic processes that give rise to the factive presupposition in the first place. In the remainder of this paper, we treat all emotives—factive (e.g. *love*, *hate*) or non-factive (e.g. *hope*, *worry*)—as both representational and preferential, since we believe it to be the most consistent treatment for our purposes. We return to this in our analysis in Section 4.

The factivity distinction appears to be tracked most closely by whether the verb allows both question and nonquestion subordinate clauses (Hintikka, 1975; Ginzburg, 1995; Lahiri, 2002; Sæbø, 2007; Egré, 2008; Uegaki, 2012; Anand and Hacquard, 2014; Spector and Egré, 2015). For instance, the factive *know* can occur with both nonquestion (15a) and question (15b) subordinate clauses,

while the nonfactive *think* can occur with nonquestion subordinate clauses (16a) but not question subordinate clauses (16b).⁶

- (15) a. Jo knows that Bo went to the store.
b. Jo knows {if, why} Bo went to the store.
- (16) a. Jo thinks that Bo went to the store.
b. *Jo thinks {if, why} Bo went to the store.

This generalization has two well-known types of exceptions. First, many nonfactive communication predicates, such as *tell* and *say*, allow both question and nonquestion subordinate clauses; second, some mental predicates, such as *decide*, *assess*, and *evaluate*, also allow both question and nonquestion subordinate clauses.

- (17) a. Jo hasn't {told me, said} whether Bo went to the store.
b. Jo hasn't yet {decided, assessed, evaluated} whether to go to the store.

Little work has been done on how learners might use the syntax to learn factives or even when factivity is acquired at all (though see Dudley et al. 2015). There is, however, some work focusing on distinctions in speaker certainty signaled by, e.g., *know* v. *think* (Harris, 1975; Johnson and Maratsos, 1977; Abbeduto and Rosenberg, 1985; Moore and Davidge, 1989; Moore et al., 1989; Schwanenflugel et al., 1994, 1996), which is likely correlated with factivity but which is a conceptually distinct phenomenon (though see Hopmann and Maratsos 1978; Scoville and Gordon 1980; Léger 2007 for tasks that attempt to test factivity).

2.3 Assertivity

Further cross-cutting representationality and factivity is the assertivity distinction (Hooper, 1975). Like factivity, assertivity is defined in terms of its effects on discourse. Again very roughly, a verb is assertive if it can be used in situations where its subordinate clause seems to carry the main point of the utterance (see Urmson 1952; Simons 2007; Anand and Hacquard 2014 for discussion). For instance, *think* and *say* seem to allow this (18a), but *hate* does not (18b).

- (18) a. **A:** Where is Jo?
B: Bo {thinks, said} that she's in Florida.
- b. **A:** Where is Jo?
B: # Bo loves that she's in Florida.

Assertivity tends to correlate with representationality, though not all representationals are assertive. For example, negative representationals (e.g. *deny*, *doubt*), fiction representationals (e.g. *imagine*, *pretend*) and emotive factives (e.g., *love*, *amaze*) are all nonassertive.

- (19) **A:** Where is Jo?
B: # Bo {doubts, is pretending} that she's in Florida.

Assertivity correlates with the availability of S-lifting and the propositional anaphor object *so*. Assertives, like *think* and *say*, can occur with S-lifted subordinate clauses (20a) and *so* (21a), but *hate* cannot occur with either S-lifting (20b) or *so* (21b).

⁶This paradigm is filled out by what Lahiri (2002) calls rogatives, like *wonder* and (for some speakers) *ask*. *Wonder*, at least, takes only subordinate questions and not nonquestions.

- (20) a. She's in Florida, Bo {thought, said}.
 b. *She's in Florida, Bo hated.
- (21) a. Bo {thinks, said} so.
 b. *Bo hates so.

Hooper (1975) claims that the assertivity distinction cross-cuts the factivity distinction to give rise to a further split between semi-factives (assertive factives), like *know*, and true factives (nonassertive factives), like *love* and *hate* (see Karttunen 1971 for an early description of this distinction).⁷ Important for our purposes is that the semi-factive v. true factive distinction appears to correlate (i) with the (semantic) representationality distinction—semi-factives also tend to be cognitive factives and true factives tend to be emotive factives—and (ii) at least two sorts of syntactic distinctions.

First, semi-factives tend to allow both polar (22a) and WH (22b) questions, but true factives tend to allow only WH questions (23b), not polar questions (23a) (Karttunen, 1977).⁸

- (22) a. Jo knows if/whether Bo sliced the bread.
 b. Jo knows if/whether Bo sliced the bread.
- (23) a. *Jo {loves, hates} if/whether Bo sliced the bread.
 b. Jo {loves, hates} how Bo sliced the bread.

Guerzoni (2007) notes, however, that this correlation is not perfect, since some canonical semi-factives like *realize* resist polar questions in many contexts.

Second, semi-factives tend to allow complementizer omission (24a), but true factives tend not to (24b). This second correlation is less strong and is likely modulated by syntax: expletive subject emotive factives appear to be better with complementizer omission, particularly when they passivize (see Grimshaw 2009 for further recent discussion of complementizer omission).

- (24) a. I {know, realize} (that) Jo already went to the store.
 b. I {hate, love} *(that) Jo already went to the store.
- (25) a. It {amazed, bothered} me ???(that) Jo already went to the store.
 b. I was {amazed, bothered} ?(that) Jo already went to the store.

The idea of assertivity as a semantic property shows up in the “function codes” developed by Gelman and Shatz (1977) and Shatz et al. (1983), and it is also important in Diessel and Tomasello’s (2001) study of parenthetical uses of propositional attitude verbs. It is also what Lewis (2013) and Lewis et al. (under revision) argue is responsible for children’s tendency to reject *think* sentences that report false beliefs: children tend to over-assume assertive parenthetical uses, and thus reject *think* sentences which they believe are used to indirectly assert something false.

2.4 Communicativity

Communicativity roughly corresponds to whether a verb refers to a communicative act. This distinction cross-cuts at least the representationality distinction—there are both representational communicatives, like *say* and *tell*, and preferential communicatives, like *demand*—and perhaps

⁷The pragmatic effects that distinguish semi-factivity from true factivity are beyond the scope of this paper. Much ink has been spilled regarding the nature of semi-factivity in recent years, however, so the interested reader is encouraged to see, e.g., Simons 2001; Abusch 2002; Abbott 2006; Romoli 2011.

⁸The traditional description of this correlation concerns cognitive v. emotive factives and not necessarily semi- v. true factives. At least in the verbal domain (and possibly outside it), these two descriptions seem to be extensionally equivalent, though which distinction is relevant does matter theoretically (cf. Abels, 2004).

other distinctions as well, such as the factive-nonfactive distinction (see Anand and Hacquard 2014 for extensive discussion of whether communicativity truly cross-cuts factivity or not).⁹

The syntactic correlates of communicativity seem quite apparent on the surface. Communicative verbs, along with a subordinate clause, tend to take noun phrase (26a) or prepositional phrase (26b) arguments representing the recipient of the communication (Zwicky, 1971).

- (26) a. Bo told me that Jo went to the store.
b. Bo said to me that Jo went to the store.

But though this is often treated as a clearly marked distinction, there are various reasons to be cautious about it. For instance, note that *demand* and *tell* can occur in string-identical contexts with *want* and *believe*. These string-identical contexts appear to be distinguished only given some parse of the string. Wanting and believing don't seem to involve anything besides a wanter/believer and a thing wanted/believed. In contrast, telling and demanding seem to require an additional role.

- (27) Bo {told, demanded, wanted, believed} Jo to be happy.

This is plausibly syntactically encoded. Note that the pleonastic element *there*, which is plausibly an overt cue to the particular syntactic configuration in question, is only allowed with *want* and *believe*, but not *tell* and *demand*. This has been used to suggest that *tell* and *demand* in (27) involve an underlying object while *want* and *believe* do not.

- (28) Bo {*told, *demanded, wanted, believed} there to be a raucous party.

Further, there are some string-identical contexts that both communicative and noncommunicative verbs can appear in which plausibly have no syntactic (or perhaps even selectional) distinctions. For instance, the communicative verb *promise* and the verb *deny*, which is plausibly noncommunicative in this syntactic context, both allow constructions with two noun phrases.

- (29) Bo {promised, denied} Bo a meal.

This is not to say that the semantic distinction has no syntactic correlates, of course; it is just to say that they may not be apparent from the string context.

2.5 Perception

The final class we consider is the perception predicates, like *see*, *hear*, and *feel*. These predicates form a somewhat small class, and all allow bare verb phrase subordinate clauses (30a) and participial verb phrase subordinate clauses (30b).

- (30) a. Jo {saw, heard, felt} Bo leave.
b. Jo {saw, heard, felt} Bo leaving.

Perception verbs were a main focus of early work on syntactic bootstrapping for verbs (Landau and Gleitman, 1985; Gleitman, 1990).

While bare verb phrase subordinate clauses are relatively distinctive of perception predicates, bare and participial verb phrase subordinate clauses can occur with predicates that do not clearly

⁹Whether *say* and *tell* are only representational is a question. Both can be used to talk about commands conditional on their taking a nonfinite subordinate clause. In any case, they plausibly have something like a representational use with finite subordinate clause.

involve perception, such as *make* and *remember*.

- (31) a. Jo made Bill leave.
b. Jo remembered Bo leaving.

2.6 Discussion

There are two main take-aways from this section. The first is that there are some potentially promising correlations between propositional attitude verbs' semantic properties and their syntactic distributions, only some of which has been studied in any depth in the acquisition literature. The second take-away is that, while these correlations are promising, it remains unclear whether they are really robust enough to support learning. This lack of clarity arises in large part because none of the correlations between particular syntactic structures and particular semantic properties are perfect, and it is difficult to assess how much this correlation improves when considering multiple syntactic structures at once without quantitative tests. In the next section, we conduct such a quantitative test.

3 Validating previous classifications

In this section, we present an experiment aimed at measuring the acceptability of a variety of propositional attitude verbs in different syntactic contexts. Our goal here is to assess the extent to which claims from traditional distributional analysis regarding correlations between syntax and semantics hold up. To do this, we compare the results of the acceptability judgment experiment against the classification of verbs discussed in the last section. This allows us to quantitatively assess how closely these previous classifications are tracked by the syntax. All experimental materials, data, and analysis code are available on the first author's github.

The methodology we use is based on one developed by Fisher et al. (1991), who were concerned that, in standard distributional analysis "...only those semantic generalizations that can be readily labeled by the investigator are likely to be discerned," but that it "...may well be that there are semantic abstractions which, while correlated with the syntax, are not so easy to puzzle out and name."

To address this concern, Fisher et al. obtain (i) a measure of verbs' meaning in the form of semantic similarity judgments and (ii) a measure of verbs' syntactic distributions, using an acceptability judgment task.¹⁰ They then ask to what extent the two measures are correlated. The idea here is that such quantitative representations allow one to bypass the sort of explicit labeling inherent to the traditional method, since distinctions among features salient to the participants are not explicitly invoked.

Fisher et al. use this methodology to study the high-level correlations between semantics and syntax, selecting a relatively small number of verbs from across the entire lexicon and using a fairly coarse-grained notion of syntactic frame. We further diverge from Fisher et al. in explicitly quantifying the relationship between the syntactic measure and prior semantic classifications, in-

¹⁰Lederer et al. (1995) took a similar tack, using the same sort of semantic similarity judgment task but replacing acceptability judgments with syntactic distributions extracted from a corpus. An anonymous reviewer asks why we did not take such a tack here. Our aim here is similar to Fisher et al.'s in that we would like to assess how much information exists in syntactic distributions *in principle*, which in turn helps us to understand which particular pieces of the syntax are associated with particular semantic properties. In other work, some of which we discuss in Section 5, we ask whether this information exists in corpora (see White 2015; White et al. under review for more details).

interpreting the relationship between the quantitative measure of the syntax and the quantitative measure of the semantics relative to these prior classifications.

3.1 Design

Thirty propositional attitude verbs were selected in such a way that they evenly spanned the classes in Hacquard and Wellwood's (2012) semantic classification. This classification is essentially a more elaborated version of the classification presented in Section 2, synthesizing much of the previous theoretical literature on the propositional attitude verb classes.¹¹

We then selected 19 syntactic features based on the theoretical literature discussed in Section 2. These features consist in five broad types: clausal complement features, noun phrase (NP) complements, prepositional phrase (PP) complements, expletive arguments, and anaphoric arguments.¹²

3.1.1 Features of interest

Six types of clausal complement features were selected: finiteness, complementizer overtness, subordinate subject overtness, subordinate question type, S-lifting, and small clause type. Finiteness had two values: finite (32a) and nonfinite (32b).

- (32) a. Jo thought that Bo went to the store.
b. Jo wanted Bo to go to the store.

Complementizer presence had two values: present (33a) and absent (33b).

- (33) a. Jo thought that Bo went to the store.
b. Jo thought Bo went to the store.

Embedded subject presence had two values: present (34a) and absent (34b) and is relevant only when the clause is finite and has no overt complementizer.

- (34) a. Jo wanted Bo to go to the store.
b. Jo wanted to go to the store.

Embedded question type had three values: nonquestion (35a), polar question (35b), and WH question (35c). Only adjunct questions were used, since constituent questions are ambiguous on the surface between a question and a free relative reading.

- (35) a. Jo knows that Bo went to the store.
b. Jo knows if Bo went to the store.
c. Jo knows why Bo went to the store.

S-lifting had two values: first person (36a) and third person (36b).

- (36) a. Bo went to the store, I think.
b. Bo went to the store, Jo said.

¹¹Other large scale attitude verb classifications exist—see, for instance, the extensions to VerbNet (Kipper-Schuler, 2005) proposed in Korhonen and Briscoe (2004); Kipper et al. (2006) and the classifications given in FrameNet (Baker et al., 1998). This classification was chosen because it hews most closely to classes discussed above and in the theoretical literature more generally.

¹²A sixth feature—degree modification—was also selected for investigation, from which we constructed four frames. We exclude this from our analyses since it was pointed out to us that the information degree modification carries is likely purely—or at least mostly—semantic in nature.

Small clause type had two values: bare small clause (37a) and gerundive small clause (37b).

- (37) a. Jo saw Bo go to the store.
b. Jo remembered going to the store.

Two NP structures were selected: single (38a) and double objects (38b). NPs were chosen so as not to have an interpretation in which they could be interpreted to have propositional content (Moulton, 2009a,b; Uegaki, 2012; Rawlins, 2013; Anand and Hacquard, 2014).

- (38) a. Jo wanted a meal.
b. Jo promised Bo a meal.

A third feature relevant to NP complements—passivization—was also included (39). The availability of structures like (39) and the unavailability of structures like (34a), appears to correlate with whether a predicate is eventive and/or encodes something about the manner in which a communicative act was performed—e.g., *say* does not encode manner while *yell* does (Postal, 1974, 1993; Pesetsky, 1991; Moulton, 2009a,b, see also Zwicky 1971 for other syntactic and semantic features that track manner of speech).

- (39) Bo was said to be intelligent.

Two types of PP complement were selected: PPs headed by *about* (40a) and PPs headed by *to* (40b).

- (40) a. Jo thought about Bo.
b. Jo said to Bo that she was happy.

Three types of expletive arguments were selected: expletive *it* matrix subject, expletive *it* matrix object, and expletive *there* matrix object/embedded subject.

- (41) a. It amazed Bo that Jo was so intelligent.¹³
b. Bo believed it that Jo was top of her class.
c. Bo wanted there to be food on the table.

Three types of anaphoric complement features were selected: *so* (42a), null complement/intransitive (42b), and nonfinite ellipsis (42c).

- (42) a. Jo knew so.
b. Jo remembered.
c. Jo wanted to.

Note for this last feature that we cannot be sure that the structure in (42b) involves null complements. See Hooper 1975; Hankamer and Sag 1976; Grimshaw 1979; Depiante 2000; Williams 2012, 2015 for further discussion of these structures.

3.1.2 Stimulus construction

These 19 features were then combined into 30 distinct abstract frames. These abstract frames are listed along the *y*-axis in Figure 1. Each categorial symbol in the frame should be interpreted as follows:

¹³It is difficult to force the subject in a sentence like (41a) to be interpreted nonreferentially. As we see in Figure 1, this likely affected the judgments for verbs like *tell*, which are fine in this frame if the subject is interpreted referentially—e.g., if *it* refers to a repository of information, such as a book.

- NP** NP constituent (e.g. *Jo*)
- WH** (Adjunct) WH word (e.g. *why*)
- V** Bare form of verb (e.g. *think*)
- VP** Verb phrase with verb in bare form (e.g. *fit the part*)
- S** Tensed clause without complementizer (e.g. *Bo fit the part*)

For each abstract frame, three instantiations were generated by inserting lexical items, resulting in 90 frame instantiations. These 90 frame instantiations were then crossed with the 30 verbs to create 3060 total items.

Thirty lists of 102 items each were then constructed subject to the restriction that the list should contain exactly 3 instances of each verb and exactly 3 instances of each frame and that the same verb should never be paired with the same frame twice in the list.¹⁴ (That is, no verb showed up with more than one instantiation of the same frame in a single list.)

These lists were then inserted into an Ibex (version 0.3-beta17) experiment script with each sentence displayed using an unmodified `AcceptabilityJudgment` controller (Drummond, 2014). This controller displays the sentence above a discrete scale. Participants can use this scale either by typing the associated number on their keyboard or by clicking the number on the scale. A 1-to-7 scale was used with endpoints labeled *awful* (1) and *perfect* (7).

3.2 Participants

Ninety participants (48 females; age: 34.2 [mean], 30.5 [median], 18–68 [range]) were recruited through Amazon Mechanical Turk (AMT) using a standard Human Intelligence Task (HIT) template designed for externally hosted experiments and modified for the specific task. Prior to viewing the HIT, participants were required to score seven or better on a nine question qualification test assessing whether they were a native speaker of American English. Along with this qualification test, participants' IP addresses were required to be associated with a location within the United States, and their HIT acceptance rates were required to be 95% or better. After finishing the experiment, participants received a 15-digit hex code, which they were instructed to enter into the HIT. Once this submission was received, participants were paid for their effort.

3.3 Data validation

Even with the stringent requirements listed above—a qualification test, IP restriction, and high HIT acceptance rate—some participants attempt to game the system. There are two main ways that participants do this: (i) submitting multiple HITs despite being instructed not to and (ii) not actually doing the task—e.g. choosing responses randomly.

The first is easy to detect. When data are submitted in Ibex, the submitting participant's IP address is converted into an MD5 hash, which is in turn associated with the responses they submit. This hash can then be used to check whether participants followed instructions in only submitting a single HIT. Two participants submitted multiple HITs: one participant submitted three and another submitted two. In both of these cases, only the first submission was used.¹⁵

The second requires more care to detect. Here, we use the fact that multiple participants did the same list. The idea is to compare each participant's responses against those of all other participants

¹⁴These lists were 102 items instead of 90 items because we are excluding four degree modification abstract frames here (see footnote 12).

¹⁵Note that this method does not distinguish between one participant attempting to submit multiple HITs from the same IP and two participants each submitting a single HIT from the same IP. We err on the side of caution in filtering all the but the first HIT from the same IP.

CUTPOINTS	ADDITIVE	MULTIPLICATIVE	LL	AIC	BIC
Equidistant	True	False	-8078	24254	52411
Equidistant	False	True	-8509	25116	53273
Equidistant	True	True	-7268	22808	51570
Varying	True	False	-7989	24084	52269
Varying	False	True	-8061	24228	52413
Varying	True	True	-7084	22448	51238

Table 1: Comparison of normalization models for acceptability judgments.

that saw the same list. If a participant has low agreement with the other participants that saw the same list and the other participants show high agreement with each other, then we conclude that the disagreeing participant was providing lower quality data and remove them from the analysis.

To implement this, we calculated Spearman rank correlations between each participant’s responses and those of every other participant that did the same list. For instance, if participants x , y , and z all did list 1, we would compute the correlation between x ’s and y ’s responses, x ’s and z ’s, and y ’s and z ’s. We then inspected the distribution of these correlations for outliers.

The median Spearman rank correlation between participant responses is 0.64 (mean=0.63, IQR=0.69-0.58). To find outliers, we use Tukey’s method. In Tukey’s method, an observation—here, a correlation between two subjects’ judgments—is deemed an outlier if that observation falls below the first quartile (Q1) minus 1.5 times the interquartile range (IQR)—i.e. the difference between the third quartile and the second—or above the third quartile (Q3) plus 1.5 times the interquartile range. Four comparisons fall below $Q1-1.5*IQR$ and none fall above $Q3+1.5*IQR$. The four that fall below are due to two participants, each from a different list. Perhaps not coincidentally, those participants were also the ones that submitted multiple HITs. The remainder of the analyses exclude responses from these two participants, resulting in 86 unique participants.

After excluding these participants, the median remains the same (to two significant figures) and the mean shifts upward slightly, from 0.63 to 0.64. (This is to be expected since the mean is more sensitive to outliers.) The IQR becomes slightly smaller, and Q1 shifts slightly upward (IQR=0.69-0.59). These correlations are comparable to those reported by Fisher et al. (1991).

3.4 Results

In this section, we investigate the extent to which attitude verb classifications presented in Section 2 are predictable from our acceptability judgment data. Prior to carrying out the actual analysis, we normalize and decorrelate the acceptability judgment data. Because we are using these normalized judgments as predictors instead of dependent variables, we need a slightly more sophisticated normalization procedure than standard z -scoring—specifically, one that controls not only for subject variability in scale use but also item variability. We describe this procedure in Section 3.4.1. It is not crucial to understand this procedure, and so readers that are more interested in the analysis itself can skip to Section 3.4.2, which discusses data decorrelation, or Section 3.4.3, which discusses the analysis itself.

3.4.1 Data normalization

The standard method for normalizing ordinal scale acceptability judgements used in the psycholinguistics literature is to normalize the data by-participant—e.g. using a method such as

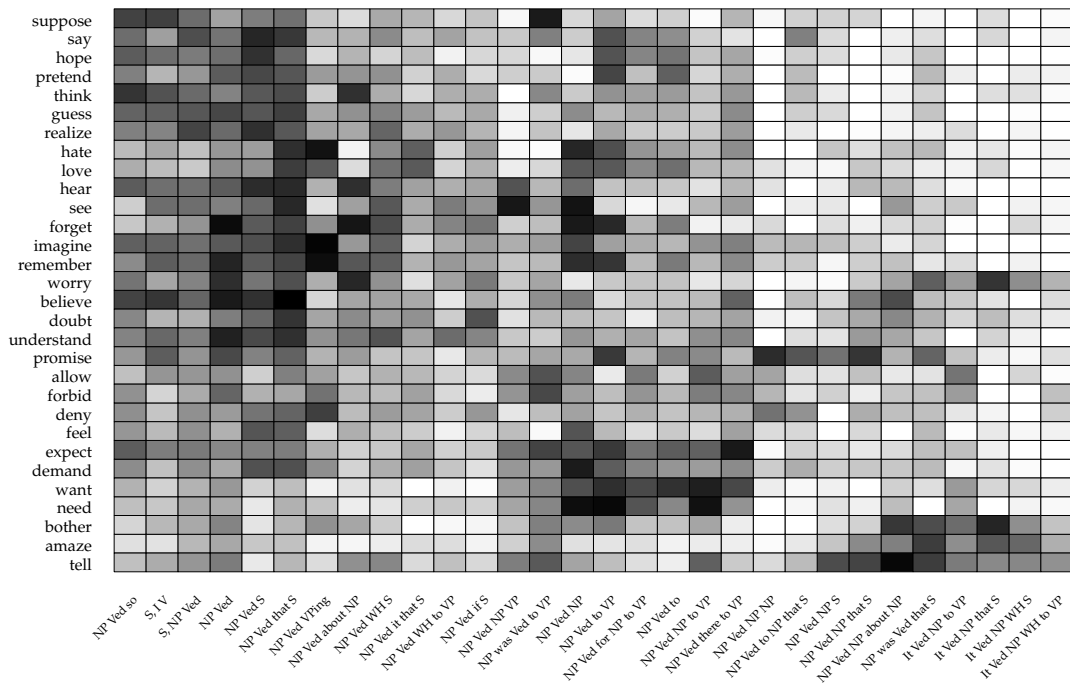


Figure 1: Normalized verb-frame acceptability judgments. Darker shades mean higher normalized ratings.

z-scoring (Schütze and Sprouse, 2014). The problem with such an approach in our case is that standard normalization methods do not control for item-based variability. Insofar as the assumptions underlying *z*-scoring are satisfied, this is not an issue in other studies, since acceptability is generally treated as a dependent variable, not a predictor, and thus item variability can be taken into account in whatever confirmatory analysis follows the transformation—generally, using random intercepts for item in a linear mixed model (Baayen et al., 2008). To address this normalization issue here, we employ an ordinal mixed model similar in form to the polytomous Rasch model (Rasch, 1960; Andersen, 1977; Andrich, 1978; Masters, 1982).

There are various ways of setting up such an ordinal mixed model that vary with respect to (i) whether ordinal ratings are associated with fixed width intervals on the normalized acceptability scale (*equidistant*) or whether those intervals can vary (*varying*), (ii) whether or not participants can vary with respect to where the midpoint of the scale lies on the normalized acceptability scale (*additive* participant effects), and (iii) whether or not participants can vary with respect to contraction of the normalized acceptability intervals (*multiplicative* participant effects). (Within this taxonomy, *z*-scoring corresponds to the equidistant model with both additive and multiplicative participant effects.)

To determine which particular normalization model to use, we employ an AIC-based model selection procedure. We fit each model using gradient descent with momentum (Rumelhart and McClelland, 1986; McClelland and Rumelhart, 1986) and learning rate annealing with a *search-then-converge* schedule (Darken and Moody, 1990) to obtain the Maximum Likelihood Estimate (MLE) for (i) the normalized acceptability of each verb-frame pair, (ii) the Best Linear Unbiased Predictors (BLUPs) for each participants along with the corresponding variance estimate, and (iii) the BLUPs for each item intercept along with the corresponding variance estimate. Each model

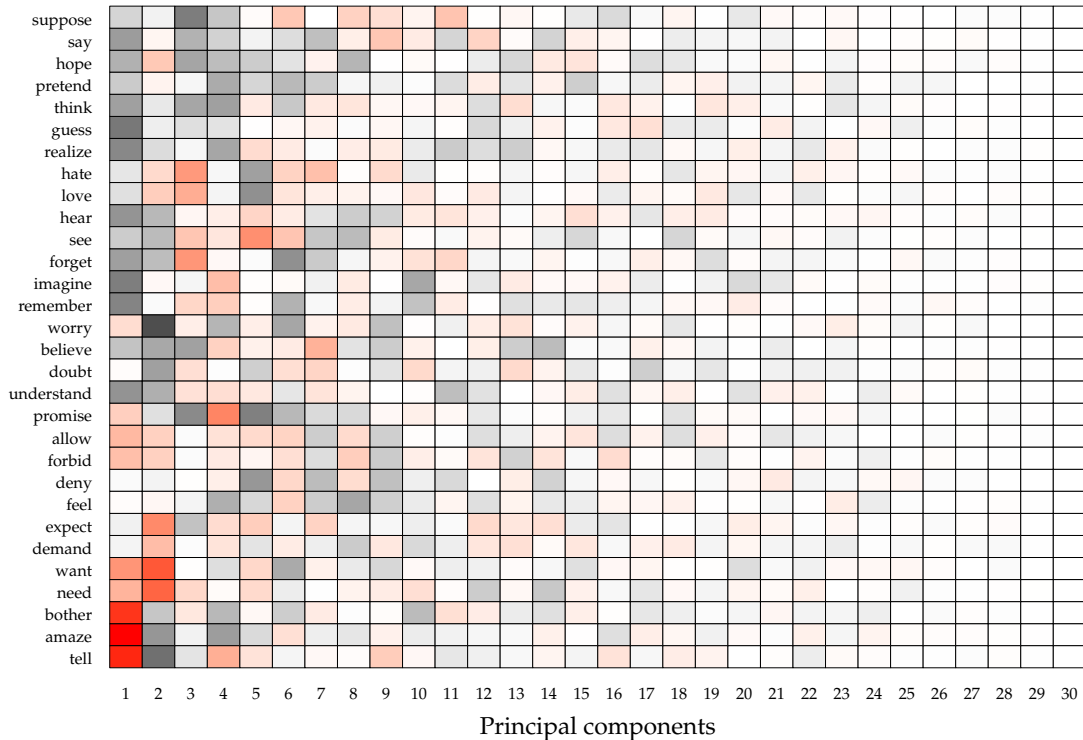


Figure 2: PCA score matrix for acceptability matrix in Figure 1.

was implemented in version 0.7 of the python package `theano` (Bergstra et al., 2010; Bastien et al., 2012).

Table 1 shows the log-likelihood, the Akaike Information Criterion (AIC; Akaike 1974), and the Bayesian Information Criterion¹⁶ (BIC; Schwarz 1978) for each of the six normalization models. The best fitting model, after penalizing for complexity using either information criterion, is the varying cutpoint additive-multiplicative model.¹⁷ Figure 1 shows the MLEs for the acceptability of each verb-frame pair when using this normalization model.

3.4.2 Data decorrelation

As can be seen by the shading of each column in Figure 1, the correlations between the normalized ratings for different subcategorization frames is quite high.¹⁸ Because we would like to use these data as predictors, we decorrelate them using Principal Component Analysis (PCA; see Jolliffe 2002). We applied PCA to the matrix of normalized judgments depicted by Figure 1 with the standard preprocessing step of first centering and standardizing by column. Figure 2 shows the PCA score matrix in descending order of their eigenvalue. Black denotes positive values and red denotes negative values, with darker shades denoting higher absolute value. Note that the scores fade off as the eigenvalues get smaller. This fading, which is expected in PCA, provides a

¹⁶We include BIC here upon reviewer request, though we follow Gelman et al. (2013) in believing BIC is not a particularly useful measure.

¹⁷This suggests that, at least for this dataset, a standard normalization such as z -scoring would have been inappropriate, even controlling for item effects. We suspect this is true for many acceptability judgment tasks, suggesting the use of z -scoring should be discouraged in favor of ordinal mixed models.

¹⁸This is not a consequence of the normalization procedure in any way. The rank correlations for the unnormalized ratings (unsurprisingly) show a nearly identical pattern.

VERB	REPR	PREF	FACT	ASSERT	COMM	PERCEPT
allow		✓				
amaze	✓	✓	✓			
believe	✓			✓		
bother	✓	✓	✓			
demand		✓			✓	
deny	✓				✓	
doubt	✓					
expect	✓			✓		
feel	✓			✓		✓
forbid		✓			✓	
forget	✓		✓	✓		
guess	✓			✓		
hate	✓	✓	✓			
hear	✓		✓	✓		✓
hope	✓	✓		✓		
imagine	✓					
love	✓	✓	✓			
need		✓				
pretend	✓					
promise	✓			✓	✓	
realize	✓		✓	✓		
remember	✓		✓	✓		
say	✓			✓	✓	
see	✓		✓	✓		✓
suppose	✓			✓		
tell	✓			✓	✓	
think	✓			✓		
understand	✓		✓	✓		
want		✓				
worry	✓	✓		✓		

Table 2: Classification of 30 verbs in experiment based on literature reviewed in Section 1.

visual cue to the importance of each component in explaining variance in the normalized ratings depicted in Figure 1.

3.4.3 Predicting attitude verb semantic properties

We now turn to our analysis of how predictable the semantic properties discussed in Section 2 are based on the normalized acceptability judgments. We consider semantic properties corresponding to the subsections of that section: REPRESENTATIONAL, PREFERENTIAL, PERCEPTION, FACTIVE, COMMUNICATIVE, and ASSERTIVE. A concise representation of which verbs have each of these properties, based on a review of the literature, is given in Table 2.

The PCA scores for each verb were entered as predictors into logistic regressions of each property.¹⁹ To determine which principal components to include as predictors, we use a step-wise

¹⁹See Rooth 1995; Stevenson and Merlo 1999; Schulte im Walde 2000; Merlo and Stevenson 2001; Korhonen 2002;

Semantic property	Accuracy
REPRESENTATIONAL	86.7
PREFERENTIAL	83.3
FACTIVE	70.0
ASSERTIVE	66.7
COMMUNICATIVE	83.3
PERCEPTION	93.3

Table 3: Cross-validation accuracy for each transformation and attitude verb class.

AIC-based model building procedure. (Using BIC does not change the final models selected.) For each semantic property, we begin with an intercept only model and allow both constructive and destructive updates of a single predictor on each step. We allow the procedure access to at most two-way interactions between principal components. As is standard in step-wise model building, interactions are only considered candidates for addition when all of their constituent predictors are already currently in the model. The only model that actually includes interactions under this procedure is the one predicting *ASSERTIVE*. Rerunning this procedure with only simple effects as candidates, does not yield a substantially worse final AIC for the *ASSERTIVE* model, and since it makes interpretation easier, we use this model instead of the one including interactions.

We find that, for every semantic property, our model-building procedure includes at least one principal component as a predictor: PC2 and PC7 for *REPRESENTATIONAL*; PC1, PC2, PC4, PC15, and PC27 for *PREFERENTIAL*; PC2, PC3, PC13, and PC27 for *FACTIVE*; PC1, PC2, PC3, PC8, PC20, and PC23 for *ASSERTIVE*; PC3, PC4, PC7, and PC12 for *COMMUNICATIVE*; and PC5 and PC8 for *PERCEPTION*. This suggests that all semantic properties are tracked at least to some extent in verbs' syntactic distributions.

As can be seen by the fact that the principal components that are chosen have low numbers, the most important principal components in terms of eigenvalue (i.e., variance explained) also tend to be the most important for predicting semantic properties. This is unsurprising, since these semantic properties are interesting exactly because they are the ones that are purported to correlate with major syntactic distinctions. But it is a useful sanity check, since if we had seen principal components with high numbers being selected, we would be suspicious that these models are fitting to noise.

Another way to make sure that we are not fitting to noise is to employ leave-one-out cross-validation for each semantic property, wherein we remove each verb from the data, train the model on the remaining data, and predict the held-out verb. We use L1 regularization as a variable selection method analogous to the step-wise procedure above.

Because L1 regularization requires us to set a regularization parameter, we use a nested leave-one-out cross-validation procedure. On each outer fold of this nested cross-validation, the PCA scores (depicted in Figure 2) and semantic properties (Table 2) for a single verb are first removed from the training set, forming the outer folds. Grid search over the L1 regularization parameter is conducted using a 4-fold crossvalidation on the resulting outer fold training set. The model selected via this grid search is then used to predict the classification of the held-out verb based on its PCA scores. This was carried out for all verbs and for all semantic properties (columns of Table 2). Table 3 shows the resulting accuracies, which are all above chance for the particular

Schulte im Walde and Brew 2002; Schulte im Walde 2003, 2006; Vlachos et al. 2008, 2009; Alishahi and Stevenson 2008 for alternative methods of evaluating existing classifications with respect to syntactic distributions.

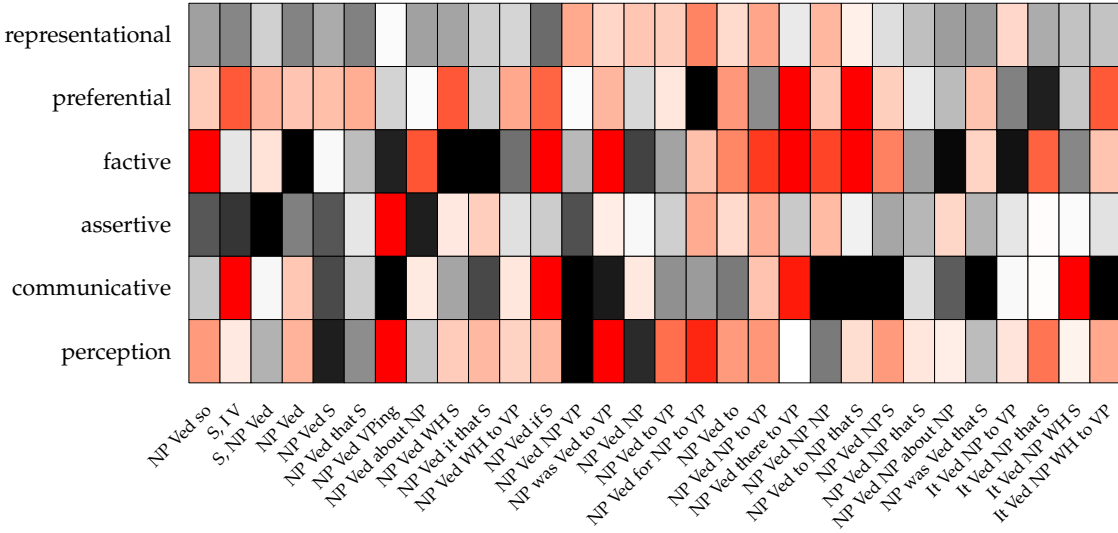


Figure 3: Frame weights for each semantic property.

semantic property. Corroborating the previous result, this suggests that all six of these semantic properties—or at least some distinction correlated with each—are tracked in the syntax.

3.4.4 Predictors of attitude verb classes

Besides knowing *that* a semantic property is predictable, it is also useful to know which frames predict it best, since these are the ones a learner might be able to use for syntactic bootstrapping. To assess this we analyze the logistic regression coefficients—which weight the principal components—in conjunction with the PCA loading matrix (not shown)—which gives the association between each principal component and each frame. We can extract the relationship between each frame and each semantic property by weighting the loading matrix by the logistic regression coefficients and summing across the latent dimensions—i.e., multiplying the vector of coefficients by the PCA loading matrix. Figure 3 shows the resulting weights for each semantic property and each frame. As for Figure 2, black denotes positive values and red denotes negative values, with darker shades denoting higher absolute value.

Figure 3 is somewhat hard to parse on its own, so to help with interpretation, we compute the rank (Spearman) correlation between the weights depicted in this figure and a dummy coding of each frame’s syntactic features, laid out in Section 3.1.1. Figure 4 shows all such correlations whose 95% confidence interval does not include zero.²⁰

To a large extent, these correlations corroborate the syntax-semantics relationships laid out in Section 2. Representationality tends to correlate with tensed clauses and propositional anaphors and allow first person S-lifting. It is also positively correlated with null complements (or intransitivity) and anticorrelated with lack of clausal tense. Preferentiality tends to show fewer such correlations overall, but it crucially does not correlate with tensed complements and what clausal complements it does correlate with are untensed. Factivity correlates with taking WH question subordinate clauses, which is consonant with the observation that factives tend to take both question and nonquestion complements. Assertivity nearly perfectly matches the distribution sug-

²⁰These confidence intervals are computed from a nonparametric bootstrap, resampling by frame, with 1,000 iterations.

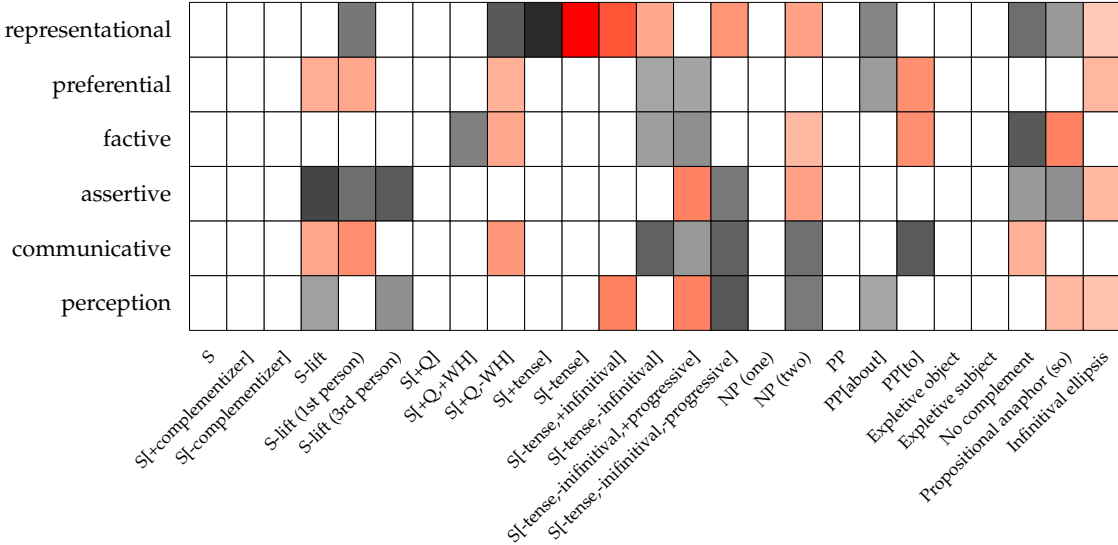


Figure 4: Feature correlations for each semantic property.

gested in the literature in correlating with S-lifting and propositional anaphors. Communicativity correlates with PP[to] complements, which tend to denote recipient arguments. And perception tends to correlate with bare verb clausal complements.

3.5 Discussion

In this section, we presented an experiment aimed at measuring the acceptability of a variety of propositional attitude verbs in different syntactic contexts. We showed that the classification from the theoretical literature, discussed in Section 2, is indeed tracked in the syntax by many of the syntactic features that are purported in the literature to track these distinctions.

But given that the accuracies in Table 3 aren't perfect, to what extent are they high enough to robustly support learning? There is no straightforward answer to this based just on these results. What we ourselves take away is that even relatively uninteresting models such as logistic regression, which have heavy constraints on the sorts of classification structures they can learn, can detect these features at least to some extent, suggesting that a more sophisticated classifier would do even better. We take it that language learners no doubt instantiate more sophisticated classifiers, and so the accuracies in Table 3 should really be seen as a strong baseline against which to test learning models.

One direction we believe would be fruitful to theoreticians and acquisitionists alike is to investigate how to account for correlations among the different semantic properties in the model. Our current setup predicts each property separately, meaning that the models we use cannot benefit from information about the space of semantic property configuration they are predicting. If the learner either innately knows or can somehow learn constraints on this space, it could make the classification task significantly easier. We leave this for future work—using, e.g., *structured prediction* methods like *conditional random fields* (Lafferty et al., 2001)—since our goal here is not the development of a learning model, but rather the establishment of baseline results.

A major question that remains at this point is to what extent the semantic properties we investigated in this section exhaust the semantic features tracked by the syntax. As noted at the beginning of this section, this question provides the original impetus for developing the method-

ology we use here (Fisher et al., 1991), but linguists have a long-standing interest in how best to approach this question—an interest that Zwicky (1971) distills quite elegantly in the introduction to his classic squib on manner-of-speech verbs.

To what extent is it possible to predict certain properties of words (syntactic, semantic, or phonological), given others? [And] insofar as there are such dependencies among properties, what general principles explain them? (ibid., p. 223)

Indeed, it has long been recognized that questions regarding which semantic distinctions are morphosyntactically (grammatically) relevant are the only ones linguists can claim propriety over; distinctions in meaning beyond those predictable from other linguistic properties fall equally well into the domain of the lexicographer (Fillmore, 1970)—or in modern times, the computer scientist (Kilgarriff, 1997). Embedded in this view is the idea that a lexical item’s linguistic contexts are responsive to only some conceivable contrasts in meaning and that a linguistic theory of the link should speak to exactly which these are and why other conceivable contrasts are excluded (cf. Jackendoff, 1972; Grimshaw, 1979; Pinker, 1989; Levin, 1993). As Zwicky puts it, the question for the linguist is “what sorts of word classes are there, and why these and not others?” (ibid., p. 223)

An example of the distinction between grammatically relevant and linguistically irrelevant semantic distinctions comes from Pesetsky (1991). Following Zwicky, he notes that, though “verbs of manner of speaking”—e.g. *holler* and *whisper*—and “verbs of content of speaking”—e.g. *say* and *propose*—are distributionally distinguishable, “verbs of loud speech”—e.g. *holler* and *shout*—and “verbs of soft speech”—e.g. *whisper* and *murmur*—do not seem to be. (For example, verbs of content of speaking “resist adjunct extraction and allow complementizer deletion” (Pesetsky, 1991, p. 14).) That is, the manner-content distinction has consequences for the syntax, whereas the loud-soft contrast does not.

In the remainder of the paper, we pivot to investigating what other semantic properties, beyond those discussed so far, might be latent in the syntax. To do this, we gather similarity judgments for each of the verbs tested in this section and ask to what extent the semantic properties from this section statistically mediate the relationship between these similarities and our normalized acceptability judgments.

4 Exhausting the semantic information

In this section, we present two experiments aimed at getting a measure of how similar in meaning naïve speakers take the propositional attitude verbs from Experiment 1 to be. Our goal here is to assess the extent to which the semantic properties discussed in Sections 2 and 3 exhaust the space of semantic properties tracked by the syntax. We use two experiments here, since as we show, different tasks seem to tap different aspects of the semantics. We begin by describing the design, methodology, validation, and normalization procedures for each experiment, and then we move onto the analysis. As for the acceptability judgment experiment presented in Section 3, all materials, data, and analysis code are available on the first author’s github.

4.1 Experiment 2a: generalized semantic discrimination task

In this first experiment, we employ a generalized semantic discrimination task—also known as a triad or “odd man out” task—in which participants are given lists of three words and asked to choose the one least like the others in meaning (Wexler, 1970; Fisher et al., 1991).

4.1.1 Design

We constructed a list containing every three-combination of the 30 verbs from Experiment 1 (4060 three-combinations total). Twenty lists of 203 items each were then constructed by randomly sampling these three-combinations, which we refer to as triads, without replacement. These lists were then inserted into an Ibex (version 0.3-beta15) experiment script with each triad displayed using an unmodified `Question` controller (Drummond, 2014). This controller displays an optional question above a list of answers. In this case, the question was omitted and the verbs making up each triad constituted the possible answers. Participants could select an answer either by typing the number associated with each answer or clicking on the answer.

4.1.2 Participants

Sixty participants (28 females; age: 34.5 [mean], 31 [median], 18–68 [range]) were recruited through AMT using a standard HIT template designed for externally hosted experiments and modified for the specific task. All qualification requirements were the same as those described in Section 3. After finishing the experiment, participants received a 15-digit hex code, which they were instructed to enter into the HIT. Once this submission was received, participants were paid for their time.

4.1.3 Data validation

The data validation procedure is the same one described in Section 3 with the exception that we calculate Cohen’s κ instead of Spearman’s ρ .²¹ The median Cohen’s κ between participant responses is 0.45 (mean=0.45, IQR=0.52-0.37).²² To find outliers, we use Tukey’s method. No comparisons fall below $Q1-1.5*IQR$ and none fall above $Q3+1.5*IQR$. Thus, we exclude no participants.

The median agreement here is quite a bit lower than the interrater agreement found by either Fisher et al. or Lederer et al.. Fisher et al. report Spearman’s $\rho=0.81$ (Exp. 1); 0.78 (Exp. 2); 0.76 (Exp. 3), 0.79 (Exp. 4), 0.72 (Exp. 5). Lederer et al. report Spearman’s $\rho=0.81$. This is likely driven by the fact that we are investigating a much smaller portion of the lexicon and thus are bound to find that participants have less certainty about which verbs are more semantically similar.²³ Another possible contributor to this lower correlation is that Cohen’s κ is more conservative than Spearman’s ρ . As we see in Section 4.2.3, however, the conservativeness of Cohen’s κ is not likely to be the culprit here, since even Spearman’s ρ shows roughly the same amount of agreement among participants using a different measure.

4.1.4 Data normalization

The fact that verbs are displayed in a list raises the worry that effects of position may arise, either as an overall preference for a particular position and/or as a participant-specific preference. We

²¹Both Fisher et al. and Lederer et al. compute Spearman rank correlations over count matrices constructed from judgments across participants. The method they use is not available to us without significant alteration since we collected data from more than two participants per list. Instead, we opt for a more standard measure of interrater agreement here. This measure is preferable in any case since (i) it allows us to assess each participant’s reliability at the same time as we assess overall agreement and (ii) it can be applied to the raw data instead of a statistic of the data, as in the cases of Fisher et al. and Lederer et al..

²²An analysis of the distribution of Fleiss’ κ (the multi-rater generalization of Scott’s d) by list corroborates this analysis (median=0.45, mean=0.45, IQR=0.48-0.40).

²³If this is indeed true, interrater agreement on this and other similarity judgment tasks could be a way of investigating the “semantic density” of a lexical neighborhood. Modeling reaction time, as a proxy for uncertainty, might also be fruitful in future research.

CUTPOINTS	ADDITIVE	MULTIPLICATIVE	LL	AIC	BIC
Equidistant	True	False	-4397	9816	12987
Equidistant	False	True	-4652	10326	13497
Equidistant	True	True	-4125	9392	12935
Varying	True	False	-4361	9752	12947
Varying	False	True	-4392	9814	13009
Varying	True	True	-4070	9290	12858

Table 4: Comparison of normalization models for ordinal similarity judgments.

see both such preferences. Across participants, there is a bias for earlier positions—proportion for position 1: 0.36, position 2: 0.34, position 3: 0.30—but substantial variability among participants—interquartile range of participant bias for position 1: [0.33, 0.39], position 2: [0.31, 0.36], position 3: [0.27, 0.34].²⁴ Thus, as in Section 3, we normalize the data prior to analysis to control for biases a particular participant may have to choose a verb in a particular position.

To carry out this normalization, we use a multinomial logistic mixed effects model. This model predicts which verb position in a triad is chosen based on (i) the (latent) similarities between each pair in the triad and (ii) the (latent) bias each participant has to choose a verb in a particular position.²⁵ We find the Maximum Likelihood Estimate (MLE) of the similarity matrix and random effects components using gradient descent implemented in version 0.7 of the python package `theano`.

4.2 Experiment 2b: ordinal similarity

In this second experiment, we employ an ordinal scale similarity task, in which participants are asked to rate the similarity in meaning of a word pair on a 1-7 scale.

4.2.1 Design

We constructed a list containing every pair of the 30 verbs from Experiment 1 plus the verb *know* (460 pairs).²⁶ Twenty lists of 62 pairs were then constructed such that every verb was seen an equal number of times and no pair was seen twice.

These lists were then inserted into an Ibex (version 0.3.7) experiment script with each pair displayed using an unmodified `AcceptabilityJudgment` controller (Drummond, 2014). This controller displays the verb pair separated by a pipe character—e.g. *think | want*—above a discrete scale. Participants could use this scale either by typing the associated number on their keyboard or by clicking the number on the scale. A 1-to-7 scale was used with endpoints labeled *very dissimilar* (1) to *very similar* (7). To encourage them to make a symmetric similarity judgment, participants

²⁴Some of this variability may be attributable to the list each participant received. Since the lists were between-subjects and we are only concerned with controlling for the variability to normalize the data, this can safely be explained by the random effects components.

²⁵We furthermore impose a symmetry constraint on the similarity matrix. A model without this constraint was also fit but had a worse AIC than the model with the symmetry constraint. This suggests that participants are in fact making judgments based on a symmetric understanding of similarity (or at least that we do not have enough evidence to show that they are not).

²⁶*Know* was added after discussion with multiple researchers suggested it may be interesting for future uses of these data. Because we do not have data about its syntactic distribution, we do not use any ratings of any pairs including *know* in our analysis.

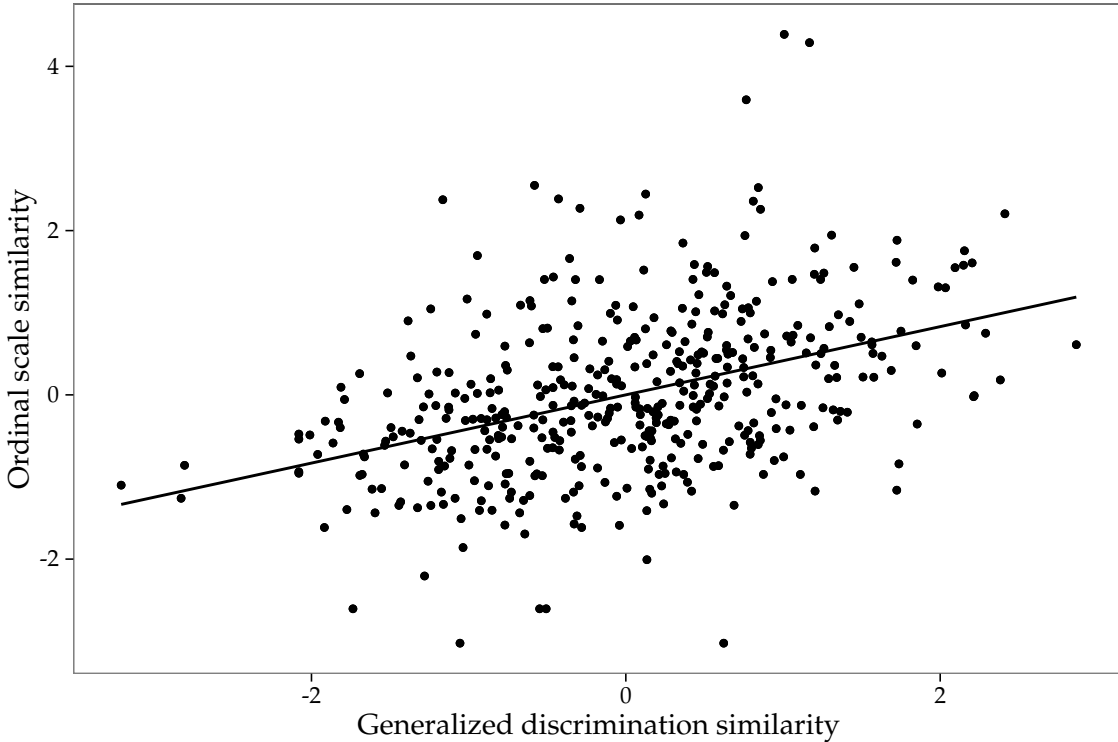


Figure 5: Generalized semantic discrimination semantic similarity ratings (normalized and standardized) plotted against ordinal scale similarity ratings (normalized and standardized).

were instructed to rate “the similarity between the meanings of the two verbs” as opposed to rating how similar the first verb was to the second (or vice versa).

4.2.2 Participants

Sixty participants were recruited through AMT. All qualification requirements were the same as those described in Section 3.2. After finishing the experiment, participants received a 15-digit hex code, which they were instructed to enter into the HIT. Once this submission was received, participants were paid for their time.

4.2.3 Data validation

The data validation procedure is the same one described in Section 3. The median Spearman rank correlation between participant responses is 0.40 (mean=0.41, IQR=0.52-0.32). To find outliers, we use Tukey’s method. No comparisons fall below $Q1-1.5*IQR$ and none fall above $Q3+1.5*IQR$. Thus, we exclude no participants.

4.2.4 Data normalization

Since these data are ordinal, we use the data normalization procedure described in Section 3.²⁷ Table 4 gives the log-likelihood and AIC for each model. As in Section 3, the best fitting model,

²⁷As for the generalized semantic discrimination normalization model, we constrain the similarities inferred to be symmetric. Each model was also fit without this constraint, and the AICs were found to be worse in comparison to the corresponding model with the constraint.

penalizing for complexity, is the model with varying cutpoints and both additive and multiplicative subject random effects. We use the MLE of the similarities inferred by this model in the remainder of this section.

4.3 Comparison of similarity datasets

Figure 5 plots the generalized semantic discrimination judgments against the ordinal scale similarity judgments, both after normalization and standardization. Overall, the correlation between responses on the generalized semantic discrimination task and those on the ordinal scale task are at about the same level as the correlations among participants within each experiment (Spearman’s $\rho=0.437$, $p < 0.001$). This suggests not only that these two tasks are tapping similar aspects of participants’ semantic knowledge but that they do so at the limit of what we would expect given inter-annotator agreement within each experiment. But as we show in the next section, the difference in the two tasks does not appear to be solely about interannotator noise; each task taps slightly different semanticoconceptual properties.

4.4 Exhausting the semantic information

If the syntax carried no information about semantic properties beyond those discussed in Sections 2 and 3, we would expect the relationship between the syntactic distributions and the similarity judgments to be mediated by those semantic properties. To assess the extent to which these semantic properties exhaust the space of semantic properties tracked by the syntax, we now conduct a mediation analysis. This analysis has three components, which we apply to each similarity dataset separately: (i) establish that there is a relationship between the semantic properties and the similarity judgments (Section 4.4.1); (ii) establish that there is a relationship between the syntactic distributions and the similarity judgments (Section 4.4.2); and (iii) measure the extent to which the relationship between the syntactic distributions and the similarity judgments remains after controlling for the relationship between the semantic properties and the similarity judgments (Section 4.4.3).

4.4.1 Predicting similarities with semantic properties

Each normalized similarity dataset was entered into a linear regression with SIMILARITY as the dependent variable and the value of REPRESENTATIONAL, PREFERENTIAL, PERCEPTION, FACTIVE, COMMUNICATIVE, and ASSERTIVE for the two verbs whose similarity is being predicted as the dependent variables. We use the same stepwise model selection procedure described in the last section, allowing up to four-way interactions.

Table 5 shows the models selected by this procedure. We don’t dwell on these models except to make three points. First, the fact that our model selection procedure does not select an intercept only model suggests that at least some of the semantic properties we have been employing correlate with naïve speakers judgments. This is important for the mediation analysis as a whole, since the mediation question would be moot if the semantic properties couldn’t in fact be mediators. Second, we see that only a subset of the semantic properties appear to be relevant in participants similarity judgments: REPRESENTATIONAL and PERCEPTION are absent from both models. Third, PREFERENTIAL only appears to be active in the generalized discrimination judgments. This is interesting because it suggests that these two tasks pick up on distinct aspects of the semantics.

	<i>Dependent variable:</i>	
	discrimination	ordinal
PREFERENTIAL	0.057 (0.122)	
FACTIVE	-0.334 (0.068)	-0.702 (0.101)
ASSERTIVE	0.641 (0.161)	-0.069 (0.094)
COMMUNICATIVE	-0.728 (0.182)	-0.793 (0.119)
PREFERENTIAL × ASSERTIVE	-0.686 (0.155)	
PREFERENTIAL × COMMUNICATIVE	0.355 (0.216)	
ASSERTIVE × COMMUNICATIVE	0.689 (0.205)	
ASSERTIVE × ASSERTIVE	-0.369 (0.176)	0.688 (0.124)
COMMUNICATIVE × COMMUNICATIVE		1.617 (0.225)
FACTIVE × FACTIVE		0.533 (0.158)
FACTIVE × COMMUNICATIVE		0.501 (0.182)
Observations	870	870
R ²	0.245	0.203
Adjusted R ²	0.232	0.193
Residual Std. Error	0.877 (df = 854)	0.898 (df = 858)
F Statistic	18.461 (df = 15; 854)	19.880 (df = 11; 858)

Table 5: Regression coefficients for predicting SIMILARITY by semantic properties

4.4.2 Predicting similarities with syntactic distributions

Each normalized similarity dataset was entered into a linear regression with SIMILARITY as the dependent variable and the principal components scores discussed in Section 3 (Figure 2) as predictors. We employ the same stepwise model selection procedure used for the semantic properties, allowing up to two-way interactions. (The space of candidate predictors explodes when allowing for larger numbers of interactions.)

The resulting models are extremely large and their parameters are not particularly easy to interpret, given that their predictors are principal components, so we do not include a table analogous to Table 5 here. It is useful, however, to compare how many simple effects are shared between the two resulting models and the semantic property model from the last section. The generalized discrimination model includes PC1, PC3, PC4, PC5, PC8, PC12, PC15, PC16, PC17, PC19, PC21, and PC28, and the ordinal scale model includes PC1, PC3, PC5, PC7, PC8, PC10, PC11, PC12, PC14, PC19, PC20, PC26, and PC28. Thus both models share PC1, PC3, PC5, PC8, PC12, PC19, and PC28—a substantial overlap in the most important principal components. There is also substantial overlap with the semantic property models. The only semantic property for which the selected model did not share any principal components was REPRESENTATIONAL, which is predicted by PC2 and PC7. The rest share at least one component.

This is interesting because it appears to further corroborate the result mentioned in the last section that REPRESENTATIONAL does not appear to be active in participants similarity judgments. This could mean that representationality is not a coherent semantic property, though if we were to accept this interpretation, we would need to explain why the syntax correlates so well with it, as we saw in Section 3.

One thing that could be happening here is that we were too promiscuous in our original coding of representationality. As mentioned in Section 2, whether emotive predicates are representational

is a contentious issue. Emotives tend to trigger inferences that involve representationality, but it is at least possible that these inferences are not semantic in nature, and thus they may not be part of emotive predicates’ semantic representations. These inferences are also typically backgrounded, and so a compounding factor may be that these inferences are not very salient to naïve speakers making similarity judgments. What we take away from this result is that one cannot rely solely on similarity judgments of the kind we use here as a source of semantic properties, even though similarity judgment tasks may be useful for large-scale annotation of some properties.

The final syntactic distribution-based models selected by the step-wise procedure show nearly double the R^2 for both the generalized discrimination judgments ($R^2=0.477$) and the ordinal scale judgments ($R^2=0.457$) compared to the semantic property-based models. This of course could be due to the fact that the syntax-based models have many more parameters. We reject this hypothesis, however, based on the fact that Vuong tests for nonnested models suggest that both the syntax-based model of the generalized discrimination judgments ($z=8.861$, $p < 0.001$) and the syntax-based model of the ordinal scale judgments ($z=7.642$, $p < 0.001$) fit significantly better than the semantic property-based models, controlling for the number of parameters.

This is suggestive that the syntactic distributions carry semantic information beyond the semantic properties we have been discussing. In the next section, we show this directly by residualizing the syntactic distributions by the semantic properties and jointly predicting the similarity judgments.

4.4.3 Exhausting the semantic information

We aim to predict the semantic similarity judgments given both the semantic properties and the syntactic distributions. We know from Section 3 that there are significant correlations between the semantic properties and the syntactic distributions, however, and so prior to carrying out this prediction, we remove this semantic property information from the syntactic distributions to avoid multicollinearity.

We fit a multivariate linear regression with the acceptability judgments for each syntactic frame as the dependent variables and semantic properties for each verb as the predictors. (Note that we use the normalized acceptability judgments here, since we are not worried about correlations among the acceptability judgments for each frame.) We then residualized the normalized acceptability judgments using this model. This results in a matrix with information about the semantic properties removed. This matrix still contains correlations among the residualized acceptability judgments for each frame, and so as in Section 3, we decorrelate these variables using PCA.

Next, we enter each normalized similarity dataset into a linear regression with SIMILARITY as the dependent variable and the principal components scores discussed in Section 3 (Figure 2) as predictors. We employ the same stepwise model selection procedure used for the semantic properties and the syntactic distributions alone. We allow up to two-way interactions between variables within each type of predictor—i.e., interactions between semantic properties and the decorrelated residualized syntactic distributions were not considered. We find that the models that this procedure selects are substantially better than the models predicting the similarity judgments based on the semantic properties alone, for both the generalized discrimination judgments ($\chi^2(29)=123.16$, $p < 0.001$) and the ordinal scale judgments ($\chi^2(26)=136.95$, $p < 0.001$). This suggests that there is further semantic information in the syntactic distributions beyond information about the semantic properties from the theoretical literature.

To assess what semantic information this is, we regress the decorrelated residualized syntactic distributions (without the semantic properties) against the similarity judgments. We then use this model to predict the similarity judgments. These predictions encode whatever semantic infor-

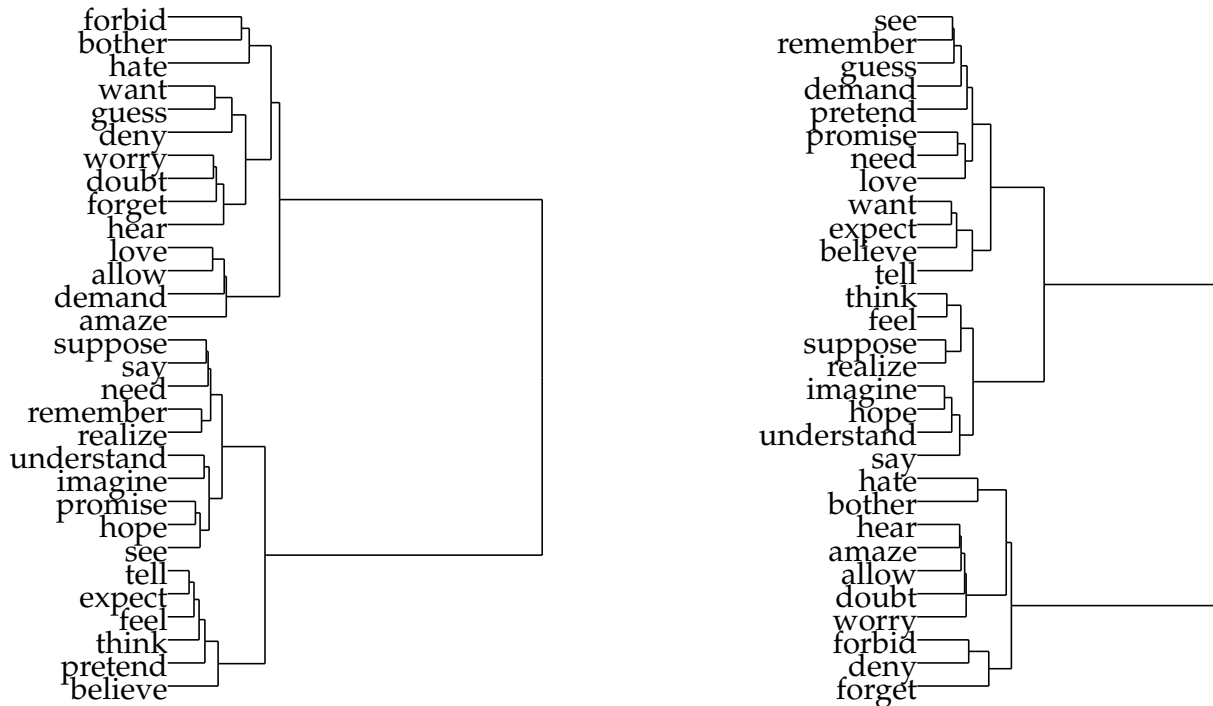


Figure 6: Hierarchical clustering of semantic information carried by the syntax about generalized discrimination judgments (left) and ordinal scale judgments (right) after removing semantic property information.

mation lies in the syntax about the particular semantic properties participants use to make their similarity judgments in the generalized discrimination and ordinal scale tasks. We again use a stepwise model-building procedure to select this model.

Figure 6 shows the results of applying a hierarchical clustering (Ward’s method) to the predicted similarity judgments for both the generalized discrimination task (left) and the ordinal scale task (right). In both cases, there is a major split between a group that contains a combination of preferentials and verbs with negative affect (whether preferential or not)—many, though not all of which, are nonassertive. For instance, the negative affect verbs *doubt*, *forget*, and *deny* are both representational nonpreferentials that occur in this cluster along with negative affect preferentials, such as *hate*, *bother*, *forbid*, and *worry*, and nonnegative affect preferentials, such as *amaze* and *allow*.

This finding is at once surprising and unsurprising from a theoretical perspective. On the one hand, many languages that have a robust mood distinction—e.g., Romance languages, like Spanish and French—group negative affect verbs together with preferentials in terms of which verbs take subjunctive subordinate clauses. (This is necessarily a rough characterization, since the distribution of subjunctive subordinate clauses turns out to be very difficult to predict precisely.) On the other hand, since English does not have the relevant distinction robustly, a property combining negative affect and preferential verbs together is not generally thought to determine syntactic distribution in English. Thus, this finding might be taken as preliminary evidence for this distinction being tracked by English syntax. Our final analysis in this section will thus be aimed at figuring out what syntactic structure this distinction correlates with.

To do this, we cut each of the trees in Figure 6 at their highest level split. We then compute the rank correlation between this split and each the normalized acceptability judgments for each frame. Figure 7 shows this rank correlation, with a positive correlation for a frame meaning that that frame correlates with the negative affect + preferential class.

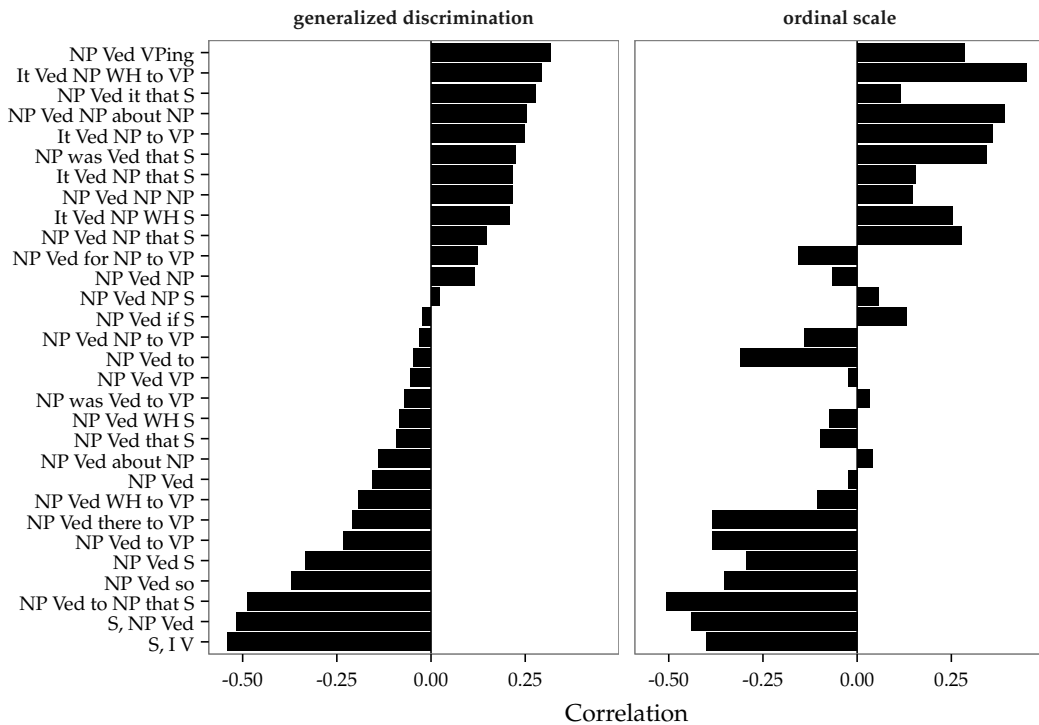


Figure 7: Correlation with negative affect + preferential class from Figure 6

What we see here is that, in general, the correlations between this class and syntactic frames is somewhat low. Rather, it is the other side of the negative affect + preferential split that seems to be tracked robustly in the syntax. Indeed, it is exactly the frames that tend to correlate with representationality and assertivity that tend to strongly correlate with this other class. What this may mean is that the negative affect + preferential class is robustly tracked in the syntax, but it is tracked in a negative sense—i.e., in terms of the syntactic structures it cannot take. This is not particularly surprising, since as we noted, many negative affect + preferential verbs also tend to be nonassertive. One possibility this finding raises is that, when a language lacks a syntactic distinction that another language uses to track a particular semantic property, that language uses alternative means of encoding that distinction distributionally—e.g., it encodes it as an elsewhere case.

4.5 Discussion

In this section, we presented two experiments aimed at getting a measure of how similar in meaning naïve speakers take the propositional attitude verbs from Experiment 1 to be. We showed that there is substantial semantic information in propositional attitude verb syntactic distributions beyond that discussed in the theoretical literature. In particular, we found that the syntax appears to track a semantic distinction that groups together preferentials and negative affect verbs, but only as an elsewhere case. This final finding raises two questions that we spend the rest of the paper discussing: (i) to what extent are the semantic properties tracked by syntactic distributions universal?; and (ii) to what extent are these relationships cross-linguistically stable?

5 General discussion

Our goal in this paper was to test the limits of syntactic bootstrapping by quantitatively assessing correlations between syntax and word meaning in the domain of propositional attitude verbs. We did this in two steps, which together amount to a mediation analysis. The first was to validate prior theoretical claims about the relationship between semantic properties and syntactic distributions. The second was to show that the semantic properties discussed in this prior work do not exhaust those tracked by the syntax.

Moving forward, we believe it will be very important to understand to what extent the correlations we find in English translate to other languages. For instance, throughout this paper we have seen that the representational-preferential distinction is quite robustly tracked by the syntax. Indeed, even the distinction related to negative affect that we discuss in Section 4 appears to be somewhat related to the representational-preferential distinction. One of the best indicators of this distinction in English is tense, corroborating claims in the attitude verbs literature. One question that arises here is how cross-linguistically stable this correlation is. The answer appears to be that it is not very stable at all; yet learners still learn these words at similar points in development (Perner et al., 2003). In the remainder of the paper, we discuss some ways that learners might overcome this instability, beginning with a discussion of what is known about cross-linguistic variability in this domain.

5.1 Unstable syntax-semantics mappings

The cross-linguistic instability in the mapping from representationality to syntactic distribution arises in at least two (but possibly more) ways: languages where the distinction is roughly tracked by mood—in the Romance languages, representationals tend to take indicative mood and preferentials tend to take subjunctive mood (Bolinger, 1968; Hooper, 1975; Farkas, 1985; Portner, 1992; Giorgi and Pianesi, 1997; Giannakidou, 1997; Quer, 1998; Villalta, 2000, 2008, a.o.)—and languages where the distinction is tracked by the availability of verb second (V2) syntax (Truckenbrodt, 2006; Scheffler, 2009).

An instance of the correlation with mood can be seen in Spanish. In Spanish both the representational (belief) verb *creer* (*think/believe*) and the preferential (desire) verb *querer* (*want*) take finite subordinate clauses. The difference between these subordinate clauses is that, whereas verbs like *creer* (*think*) take subordinate clauses with verbs inflected for indicative mood (43a), verbs like *querer* (*want*) take subordinate clauses with verbs inflected for subjunctive mood (43b).

- (43) 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.

This makes the subordinate clause under *creer* look more like the declarative main clause in Spanish, whose tensed verb is inflected for indicative mood.

- (44) Peter va a la casa.
 Peter go.PRES.IND to the house.

An instance of the correlation with V2 can be seen in German and other Germanic languages. V2, which is generally found in main clauses, is a phenomenon in which a clause's tensed verb

appears as the second word in a sentence. For instance, (45) shows a German main clause with the tensed form of the auxiliary verb *sein* (*be*) occurring as the second word of the sentence (in second position).

- (45) Peter ist nach Hause gegangen
Peter is to home gone

In subordinate clauses headed by the complementizer *dass* (*that*), this verb occurs clause-finally, which evidences the fact that German is underlyingly a subject-object-verb (SOV) language. Both the verb *glauben* (*think*) and the verb *wollen* (*want*) can take such clauses, in which the main verb is tensed.

- (46) a. Ich glaube, dass Peter nach Hause gegangen ist.
I think that Peter to home gone is.
b. Ich will, dass Peter nach Hause geht.
I want that Peter to home goes.

Only *glauben* (*think*), however, allows a second sort of structure more akin to the main clause in the position of the tensed verb (Scheffler, 2009). If the complementizer *dass* (*that*) is not present, *glauben* (*think*) can take a subordinate clause with syntax that looks exactly like that of the main clause—compare the main clause in (45) with the subordinate clause in (47a). *Wollen* does not allow this (47b).

- (47) a. Ich glaube, Peter ist nach Hause gegangen.
I think Peter is to home gone.
b. *Ich will, Peter geht nach Hause.
I want Peter goes to home.

Thus, though both Spanish and German take tensed complements, militating against a hard-coded link between tense and representationality, they still show language-internal correlations between representationality and some more abstract aspect of the clausal syntax. Further, the aspect of the clausal syntax that occurs with only the representational verbs—indicative mood in Spanish and V2 in German—also tends to show up in declarative main clauses.

5.2 The importance of main clause syntax

This apparent language-internal correlation has led some authors to conclude that, rather than there being a relationship directly between representationality and tense, as is evidenced in English, the relationship needs to be specified more abstractly. One idea is that this more abstract mapping between semantics and syntax should be specified in terms of (*declarative*) *main clause syntax* (Dayal and Grimshaw, 2009; Hacquard, 2014). One reason that such a correlation might exist is that (declarative) main clauses are often used to assert content and many representationals are assertive.

Under this view, the apparent relationship between tense in English, mood in Spanish (and the rest of Romance), and V2 in German (and other Germanic languages besides English) is really the outgrowth of a more abstract relationship between some cluster of syntactic features—call them MAIN CLAUSE features—that are language-specific but likely highly constrained. The way in which they are constrained is that they tend to be associated with properties of the subordinate clause's that are “close” to the attitude verb. For instance, both complementizers and mood have been argued to occur high within the clausal structure (cf. Cinque, 1999; Speas, 2004), which

in turn seems to make them amenable to selection by particular semantic classes of verbs—e.g. representationals or preferentials. Indeed, ideally, one could pin the relevant feature to some particular type of head which carries the relevant selection information—e.g. the complementizer—and is “as high as possible” within the subordinate clause so as to make selection maximally local.

Suggestive of this possibility is that the standard analysis of German V2, which has that V2 is a particular kind of complementizer-driven movement akin to that seen in English WH-movement (Den Besten, 1983). English may be amenable to such an analysis in the sense that complementizer drop with finite subordinate clauses tends to only occur with representationals (Dayal and Grimshaw, 2009).

- (48) a. Bo {thinks, believes, knows} (that) Jo is out of town.
b. Bo {loves, hates} *(that) Jo is out of town.

This latter fact is furthermore suggestive, since of course English main clauses do not have complementizers, bolstering the relationship between main clause syntax and representationality, at least in English. This, however, also raises a potential problem for languages like Spanish, which lack complementizer drop in any subordinate clauses but whose declarative main clauses do not have complementizers.

But regardless of whether main clause syntax information can be carried solely in the complementizers themselves—thus allowing for an extremely local form of selection giving rise to the relationship between representationality and main clause syntax—or whether somewhat longer distance relationships need to be posited, there is nonetheless a potential relationship between the representational-preferential distinction and this language-specific-yet-highly-constrained MAIN CLAUSE feature.

The importance of this for current purposes is that, if such a correlation between representational and main clause syntax exists, it may signal a possible candidate for a hard-coded-yet-flexible projection rule. This is useful because it points to a potentially cross-linguistically robust solution to the *labeling problem* in syntactic bootstrapping: once the learner knows that a particular set of verbs belong to a semantic class, how do they determine which particular semantic class that is? Further, since the main clause syntax itself is presumably observable to the same extent that subordinate clause syntax is, the language specific instantiation of the MAIN CLAUSE feature may well itself be learnable.

In current work, we are investigating what form hard-coded-yet-flexible projection rule might take in syntactic bootstrapping (White 2015; White et al. under review; see also White and Rawlins to appear for related work). The basic idea is that learners begin with access to a set of unvalued syntactic features—e.g., [+/- SUBJUNCTIVE], [+/- TENSE]—that a particular abstract structure—in this case, MAIN CLAUSE—will instantiate, along with a rule that tells them which semantic property verbs that embed clauses with features similar to that structure instantiate—in this case, REPRESENTATIONAL \leftarrow MAIN CLAUSE. They must then identify what the actual feature valuation for the abstract structure is in order to figure out how to use the rule.

In preliminary research on English, we have found that computational models of syntactic bootstrapping that instantiate this idea not only learn the correct valuation of features for MAIN CLAUSE correctly, but they do so extremely quickly—in large part because main clauses are by definition extremely prevalent in the input. We believe this strategy may be extendable beyond just declarative main clauses to, e.g., questions and imperatives, and are currently working on further modeling to determine if this is feasible.

5.3 Conclusion

Theoretical syntax and semantics literature and the language acquisition literature have largely remained separate. On the one hand, the theoretical literature has focused on understanding the fine-grained relationships that exist between word meaning and syntactic structure, without much thought to whether these relationships are actually robust enough to support learning. On the other hand, the acquisition literature has focused on how only very few syntactic distinctions are leveraged in verb learning.

In this paper, we attempted to bridge this divide by combining the sort of rigorous quantitative techniques employed in experimental and computational approaches to language acquisition with close attention to theoretical proposals about linguistically relevant properties of meaning. We believe this general approach of quantitatively assessing theoretical proposals will prove to be fruitful for both our understanding of the acquisition of word meaning and for semantic theory more generally.

References

- Abbeduto, Leonard, and Sheldon Rosenberg. 1985. Children's knowledge of the presuppositions of know and other cognitive verbs. *Journal of Child Language* 12:621–641.
- Abbott, Barbara. 2006. Where have some of the presuppositions gone. In *Drawing the Boundaries of Meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, ed. G.L. Ward and B.J. Birner, 1–20. John Benjamins Publishing Company.
- Abels, Klaus. 2004. Why "surprise"-predicates do not embed polar interrogatives. In *Linguistische Arbeitsberichte*, volume 79, 203–221. Leipzig: Universität Leipzig.
- Abusch, Dorit. 2002. Lexical alternatives as a source of pragmatic presuppositions. In *Semantics and Linguistic Theory*, volume 12, 1–19.
- Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.
- 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.
- Anand, Pranav, and Valentine Hacquard. 2014. Factivity, belief and discourse. In *The Art and Craft of Semantics: A Festschrift for Irene Heim*, ed. Luka Crnić and Uli Sauerland, volume 1, 69–90. Cambridge, MA: MIT Working Papers in Linguistics.
- Andersen, Erling B. 1977. Sufficient statistics and latent trait models. *Psychometrika* 42:69–81.
- Andrich, David. 1978. A rating formulation for ordered response categories. *Psychometrika* 43:561–573.
- Asher, Nicholas. 2000. Truth conditional discourse semantics for parentheticals. *Journal of Semantics* 17:31–50.

- Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59:390–412.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, 86–90. Association for Computational Linguistics.
- 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 36th Annual Conference of the Cognitive Science Society*.
- Barak, Libby, Afsaneh Fazly, and Suzanne Stevenson. 2014b. Learning verb classes in an incremental model. *Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics* 37–45.
- Bastien, Frédéric, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian Goodfellow, Arnaud Bergeron, Nicolas Bouchard, David Warde-Farley, and Yoshua Bengio. 2012. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*.
- Bergstra, James, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. 2010. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for scientific computing conference (SciPy)*, volume 4, 3. Austin, TX.
- Bolinger, Dwight. 1968. Postposed main phrases: an English rule for the Romance subjunctive. *Canadian Journal of Linguistics* 14:3–30.
- Brown, Roger. 1957. Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology* 55:1.
- Brown, Roger. 1973. *A first language: The early stages*. Harvard University Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.
- Darken, Christian, and John Moody. 1990. Note on learning rate schedules for stochastic optimization. In *Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems 3, NIPS-3*, 832–838. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Dayal, Veneeta, and Jane Grimshaw. 2009. Subordination at the interface: the Quasi-Subordination Hypothesis.
- De Villiers, Jill. 2007. The interface of language and theory of mind. *Lingua* 117:1858–1878.
- 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.
- Den Besten, Hans. 1983. On the interaction of root transformations and lexical deletive rules. *On the formal syntax of the Westgermania* 47–131.
- Depiante, Marcela Andrea. 2000. The syntax of deep and surface anaphora: a study of null complement anaphora and stripping/bare argument ellipsis. Doctoral Dissertation, University of Connecticut.
- Diessel, Holger, and Michael Tomasello. 2001. *The acquisition of finite complement clauses in English: A corpus-based analysis*.
- Drummond, Alex. 2014. Ibex. URL <https://github.com/addrummond/ibex>.
- Dudley, Rachel, Naho Orita, Valentine Hacquard, and Jeffrey Lidz. 2015. Three-year-olds' understanding of know and think. In *Experimental perspectives on presuppositions*, 241–262. Springer.
- Egré, Paul. 2008. Question-embedding and factivity. *Grazer Philosophische Studien* 77:85–125.
- Farkas, Donka. 1985. *Intensional descriptions and the Romance subjunctive mood*. Taylor & Francis.
- Fillmore, Charles John. 1970. The grammar of hitting and breaking. In *Readings in English Transformational Grammar*, ed. R.A. Jacobs and P.S. Rosenbaum, 120–133. Waltham, MA.: Ginn.
- Fisher, Cynthia. 1994. Structure and meaning in the verb lexicon: Input for a syntax-aided verb learning procedure. *Language and Cognitive Processes* 9:473–517.
- Fisher, Cynthia, Yael Gertner, Rose M. Scott, and Sylvia Yuan. 2010. Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science* 1:143–149.
- Fisher, Cynthia, Henry Gleitman, and Lila R. Gleitman. 1991. On the semantic content of subcategorization frames. *Cognitive Psychology* 23:331–392.
- Fisher, Cynthia, D. Geoffrey Hall, Susan Rakowitz, and Lila Gleitman. 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua* 92:333–375.
- Gelman, Andrew, Jessica Hwang, and Aki Vehtari. 2013. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 1–20.
- Gelman, Rochel, and Marilyn Shatz. 1977. Appropriate speech adjustments: The operation of conversational constraints on talk to two-year-olds. *Interaction, conversation, and the development of language*. New York: Wiley .
- Giannakidou, Anastasia. 1997. The landscape of polarity items. Doctoral Dissertation, University of Groningen.
- Gillette, Jane, Henry Gleitman, Lila Gleitman, and Anne Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73:135–176.
- Ginzburg, Jonathan. 1995. Resolving questions, II. *Linguistics and Philosophy* 18:567–609.

- Giorgi, Alessandra, and Fabio Pianesi. 1997. *Tense and Aspect: Form Semantics to Morphosyntax*. Oxford: Oxford University Press.
- Gleitman, Lila. 1990. The structural sources of verb meanings. *Language acquisition* 1:3–55.
- Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–326.
- Grimshaw, Jane. 2009. That’s nothing: the grammar of complementizer omission.
- Guerzoni, Elena. 2007. Weak Exhaustivity and ‘Whether’: A Pragmatic Approach. In *Semantics and Linguistic Theory*, volume 17, 112–129.
- Hacquard, Valentine. 2014. Bootstrapping attitudes. In *Semantics and Linguistic Theory*, volume 24, 330–352.
- Hacquard, Valentine, and Alexis Wellwood. 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and Pragmatics* 5:1–29.
- Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 391–428.
- Harrigan, Kaitlyn. 2015. Syntactic bootstrapping in the acquisition of attitude verbs. Doctoral Dissertation, University of Maryland.
- Harrigan, Kaitlyn, Valentine Hacquard, and Jeffrey Lidz. 2016. Syntactic Bootstrapping in the Acquisition of Attitude Verbs: think, want and hope. In *Proceedings of WCCFL 33*, ed. Kyeongmin Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson, and Lisa Shorten. Cascadia Proceedings Project.
- Harris, Richard J. 1975. Children’s comprehension of complex sentences. *Journal of Experimental Child Psychology* 19:420–433.
- Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of semantics* 9:183–221.
- Hintikka, Jaakko. 1975. Different Constructions in Terms of the Basic Epistemological Verbs: A Survey of Some Problems and Proposals. In *The Intentions of Intentionality and Other New Models for Modalities*, 1–25. Dordrecht: D. Reidel.
- Hooper, Joan B. 1975. On Assertive Predicates. In *Syntax and Semantics*, ed. John P. Kimball, volume 4, 91–124. New York: Academy Press.
- Hopmann, Marita R., and Michael P. Maratsos. 1978. A developmental study of factivity and negation in complex syntax. *Journal of Child Language* 5:295–309.
- Horn, Laurence Robert. 1972. On the semantic properties of logical operators in English. Doctoral Dissertation, UCLA.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Johnson, Carl Nils, and Michael P. Maratsos. 1977. Early comprehension of mental verbs: Think and know. *Child Development* 1743–1747.
- Jolliffe, Ian. 2002. *Principal Component Analysis*. John Wiley & Sons.
- Karttunen, Lauri. 1971. Some observations on factivity. *Paper in Linguistics* 4:55–69.

- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3–44.
- Kilgarriff, Adam. 1997. I don't believe in word senses. *Computers and the Humanities* 31:91–113.
- Kiparsky, Paul, and Carol Kiparsky. 1970. Fact. In *Progress in Linguistics: A Collection of Papers*, ed. Manfred Bierwisch and Karl Erich Heidolph, 143–173. The Hague: Mouton.
- Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *Proceedings of LREC*, volume 2006.
- Kipper-Schuler, Karin. 2005. VerbNet: A broad-coverage, comprehensive verb lexicon. Doctoral Dissertation, University of Pennsylvania.
- Korhonen, Anna. 2002. Subcategorization Acquisition. Doctoral Dissertation, University of Cambridge.
- Korhonen, Anna, and Ted Briscoe. 2004. Extended lexical-semantic classification of English verbs. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, 38–45. Association for Computational Linguistics.
- Lafferty, John, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth international conference on machine learning, ICML*, volume 1, 282–289.
- Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford University Press.
- Landau, Barbara, and Lila R. Gleitman. 1985. *Language and experience: Evidence from the blind child*, volume 8. Harvard University Press.
- Lederer, Anne, Henry Gleitman, and Lila Gleitman. 1995. Verbs of a feather flock together: semantic information in the structure of maternal speech. In *Beyond Names for Things: Young Children's Acquisition of Verbs*, ed. M. Tomasello and W.E. Merriman, 277–297. Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Lewis, Shevaun, Valentine Hacquard, and Jeffrey Lidz. under revision. "Think" pragmatically: Children's interpretation of belief reports.
- 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.
- Léger, Catherine. 2007. The acquisition of two types of factive complements. In *Language acquisition and development: Proceedings of GALA*, 337–347.
- Macnamara, John. 1972. Cognitive basis of language learning in infants. *Psychological review* 79:1.
- Masters, Geoff N. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47:149–174.

- McClelland, James L., and David E. Rumelhart. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Psychological and Biological Models*, volume 2. Cambridge, MA: MIT Press.
- Merlo, Paola, and Suzanne Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics* 27:373–408.
- Moore, Chris, Dana Bryant, and David Furrow. 1989. Mental terms and the development of certainty. *Child Development* 167–171.
- Moore, Chris, and Jane Davidge. 1989. The development of mental terms: Pragmatics or semantics? *Journal of child Language* 16:633–641.
- Moulton, Keir. 2009a. Clausal Complementation and the Wager-class. In *Proceedings of the 38th annual meeting of the North East Linguistic Society*, ed. Anisa Schardl, Martin Walkow, and Muhammad Abdurrahman, 165–178. Amherst, MA: GLSA.
- Moulton, Keir. 2009b. Natural selection and the syntax of clausal complementation. Doctoral Dissertation, University of Massachusetts, Amherst.
- Naigles, L., Henry Gleitman, and Lila Gleitman. 1993. Syntactic bootstrapping and verb acquisition. *Language and cognition: a developmental perspective*. Norwood, NJ: Ablex .
- Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of child language* 17:357–374.
- Naigles, Letitia. 2000. Manipulating the input: Studies in mental verb acquisition. *Perception, cognition, and language: Essays in honor of Henry and Lila Gleitman* 245–274.
- Naigles, Letitia R. 1996. The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition* 58:221–251.
- Papafragou, Anna, Kimberly Cassidy, and Lila Gleitman. 2007. When we think about thinking: The acquisition of belief verbs. *Cognition* 105:125–165.
- 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.
- Pesetsky, David. 1991. Zero syntax: vol. 2: Infinitives.
- Pinker, Steven. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
- Portner, Paul. 1992. Situation theory and the semantics of propositional expressions. Doctoral Dissertation, University of Massachusetts, Amherst.
- Portner, Paul, and Aynat Rubinstein. 2013. Mood and contextual commitment. In *Semantics and Linguistic Theory*, volume 22, 461–487.
- Postal, Paul M. 1993. Some defective paradigms. *Linguistic Inquiry* 347–364.
- Postal, Paul Martin. 1974. *On raising: one rule of English grammar and its theoretical implications*. Current Studies in Linguistics. Cambridge, MA: MIT Press.

- Quer, Josep. 1998. Mood at the Interface. Doctoral Dissertation, Utrecht Institute of Linguistics, OTS.
- Rasch, Georg. 1960. *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press.
- Rawlins, Kyle. 2013. About 'about'. In *Semantics and Linguistic Theory*, ed. Todd Snider, volume 23, 336–357.
- Reinhart, Tanya. 1983. Point of view in language—The use of parentheticals. In *Essays on Deixis*, ed. Gisa Rauh, volume 188, 169–194. Tübingen: Narr.
- Romoli, Jacopo. 2011. The presuppositions of soft triggers aren't presuppositions. In *Semantics and Linguistic Theory*, volume 21, 236–256.
- Rooryck, Johan. 2001. Evidentiality, part I. *Glott International* 5:125–133.
- Rooth, Mats. 1995. Two-dimensional clusters in grammatical relations. In *AAAI Symposium on Representation and Acquisition of Lexical Knowledge*. Stanford, CA.
- Ross, John Robert. 1973. Slifting. In *The formal analysis of natural languages*, ed. Maurice Gross, Morris Halle, and Marcel-Paul Schützenberger, 133–169. The Hague: Mouton de Gruyter.
- Rumelhart, David E., and James L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, volume 1. Cambridge, MA: MIT Press.
- Scheffler, Tatjana. 2009. Evidentiality and German attitude verbs. *University of Pennsylvania Working Papers in Linguistics* 15.
- Schwanenflugel, Paula J., William V. Fabricius, and Caroline R. Noyes. 1996. Developing organization of mental verbs: Evidence for the development of a constructivist theory of mind in middle childhood. *Cognitive Development* 11:265–294.
- Schwanenflugel, Paula J., William V. Fabricius, Caroline R. Noyes, Kelleigh D. Bigler, and Joyce M. Alexander. 1994. The organization of mental verbs and folk theories of knowing. *Journal of Memory and Language* 33:376.
- Schwarz, Gideon. 1978. Estimating the dimension of a model. *The annals of statistics* 6:461–464.
- Schütze, Carson T., and Jon Sprouse. 2014. Judgment data. In *Research Methods in Linguistics*, ed. Robert J. Podesva and Devyani Sharma, 27–50. Cambridge University Press.
- Scoville, Richard P., and Alice M. Gordon. 1980. Children's understanding of factive presuppositions: An experiment and a review. *Journal of Child Language* 7:381–399.
- Shatz, Marilyn, Henry M. Wellman, and Sharon Silber. 1983. The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition* 14:301–321.
- Simons, Mandy. 2001. On the conversational basis of some presuppositions. In *Semantics and Linguistic Theory*, ed. R. Hasting, B. Jackson, and Z. Zvolensky, volume 11, 431–448. Ithaca, NY: Cornell University.
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117:1034–1056.

- Snedeker, Jesse, and Lila Gleitman. 2004. Why it is hard to label our concepts. In *Weaving a Lexicon*, 257–294. Cambridge, MA: MIT Press.
- Speas, Margaret. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114:255–276.
- Spector, Benjamin, and Paul Egré. 2015. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese* 192:1729–1784.
- Stalnaker, Robert. 1973. Presuppositions. *Journal of philosophical logic* 2:447–457.
- Stalnaker, Robert. 1984. *Inquiry*. Cambridge University Press.
- Stevenson, Suzanne, and Paola Merlo. 1999. Automatic verb classification using distributions of grammatical features. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, 45–52. Association for Computational Linguistics.
- Sæbø, Kjell Johan. 2007. A whether forecast. In *Logic, Language, and Computation*, ed. B.D. ten Cate and H.W. Zeevat, 189–199. Verlag, Berlin, Heidelberg: Springer.
- Truckenbrodt, Hubert. 2006. On the semantic motivation of syntactic verb movement to C in German. *Theoretical Linguistics* 32:257–306.
- Uegaki, Wataru. 2012. Content nouns and the semantics of question-embedding predicates. *Proceedings of Sinn und Bedeutung* 16 .
- 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. 1995. Questioning minds and answering machines. In *Proceedings of the 19th Boston University Conference on Language Development*, 20–36.
- 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.
- Vlachos, Andreas, Zoubin Ghahramani, and Anna Korhonen. 2008. Dirichlet process mixture models for verb clustering. In *Proceedings of the ICML workshop on Prior Knowledge for Text and Language*.
- Vlachos, Andreas, Anna Korhonen, and Zoubin Ghahramani. 2009. Unsupervised and constrained Dirichlet process mixture models for verb clustering. In *Proceedings of the workshop on geometrical models of natural language semantics*, 74–82. Association for Computational Linguistics.
- 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.
- Schulte im Walde, Sabine. 2003. Experiments on the Automatic Induction of German Semantic Verb Classes. Doctoral Dissertation, Universität Stuttgart.

- Schulte im Walde, Sabine. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics* 32:159–194.
- Schulte im Walde, Sabine, and 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.
- Waxman, Sandra R., and Jeffrey L. Lidz. 2006. Early Word Learning. In *Handbook of child psychology: Vol 2, Cognition, perception, and language (6th ed.)*, ed. D. Kuhn, R. S. Siegler, W. Damon, and R. M. Lerner, 299–335. Hoboken, NJ, US: John Wiley & Sons Inc.
- Waxman, Sandra R., and Dana B. Markow. 1995. Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive psychology* 29:257–302.
- Wexler, Kenneth N. 1970. Embedding structures for semantics. University of California, Irvine.
- White, Aaron Steven. 2015. Information and Incrementality in Syntactic Bootstrapping. Doctoral Dissertation, University of Maryland.
- White, Aaron Steven, Valentine Hacquard, and Jeffrey Lidz. under review. Main clause syntax and the labeling problem in syntactic bootstrapping. In *Semantics in Acquisition*, ed. Kristen Syrett and Sudha Arunachalam, Trends in Language Acquisition Research (TiLAR). John Benjamins Publishing Company.
- White, Aaron Steven, and Kyle Rawlins. to appear. A computational model of S-selection. In *Semantics and Linguistic Theory*, volume 26.
- Williams, Alexander. 2012. Null Complement Anaphors as definite descriptions. In *Semantics and Linguistic Theory*, volume 22, 125–145.
- Williams, Alexander. 2015. *Arguments in Syntax and Semantics*. Cambridge University Press.
- Wimmer, Heinz, and Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13:103–128.
- Zwicky, Arnold M. 1971. In a manner of speaking. *Linguistic Inquiry* 2:223–233.