Title

A rational model of syntactic bootstrapping

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Gauthier, Jon
Levy, Roger P.
Tenenbaum, Joshua B.

Publication Date

2019

Peer reviewed

A rational model of syntactic bootstrapping

Jon Gauthier, Roger P. Levy, Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
jon@gauthiers.net, {rplevy, jbt}@mit.edu

Abstract

Children exploit regular links between the meanings of words and the syntactic structures in which they appear to learn about novel words. This phenomenon, known as *syntactic bootstrapping*, is thought to play a critical role in word learning, especially for words with more opaque meanings such as verbs. We present a computational word learning model which reproduces such syntactic bootstrapping phenomena after exposure to a naturalistic word learning dataset, even when under substantial memory constraints. The model demonstrates how experimental syntactic bootstrapping effects constitute rational behavior given the nature of natural language input. The model unifies computational accounts of word learning and syntactic bootstrapping effects observed in the laboratory, and offers a path forward for demonstrating the broad power of the syntax–semantics link in language acquisition.

Keywords: syntactic bootstrapping; word learning; computational models

Children face multiple challenges of induction when acquiring their first language. They must work out the most fundamental features of language: that words exist, and that they can be used to refer to entities and relations out in the world. At a higher level, they must work out what words actually mean, and how those words can productively combine with other words to form phrases and sentences.

A successful research program has identified how children as young as 13 months can learn the meanings of a particular class of words — concrete nouns — from noisy observations of adult language use (Smith and Yu, 2008; Trueswell et al., 2013). While nouns often pick out concrete referents which are easily identifiable by a listener, other classes of words pose more substantial learning problems. Verbs, for example, often have no concrete reference in the perceptual world which the child directly observes. Certain verb meanings may also be under-determined by the perceptual facts: verb pairs such as *chase* and *flee* or *hit* and *kick* often pick out the same events, though they have vastly different meanings (Gleitman et al., 2005).

These features make learning verb meanings a challenge for both children and adults. The productive vocabularies of young children are heavily skewed toward frequent nouns with concrete referents (Fenson et al., 1994). Adult subjects in laboratory language learning experiments also routinely struggle to identify verb meanings from observations of their use (Gillette et al., 1999). But children somehow climb over these learning barriers to become adults who can *give* and *take* or *hit* and *kick*. We must account, then, for how that learning goes through. First, because verbs make reference to abstract events and relations between entities, we must account for the representations of such events and relations in the mind of the child. In other words, we must account for the **target** representations of word learning. Second, we must explain what information **sources** children exploit in order to learn which words pick out which events and relations. Because perceptual information under-determines the solution to this learning problem, there must be other sources of information in the learner's experience which help determine the meanings of these words.

This paper addresses the theory of *syntactic bootstrapping*, which claims that children exploit systematic relations between the syntactic structures in which verbs are used and their semantics in order to learn about the meanings of novel words (Landau and Gleitman, 1985; Fisher et al., 2010). After reviewing corpus and experimental evidence regarding the syntax–semantics link, we formalize syntactic bootstrapping in a probabilistic computational model, proceeding from minimal assumptions about the structure of the lexicon to a model which replicates the qualitative behavior of children in syntactic bootstrapping experiments. We show how the knowledge assumed by this model can be learned *from scratch* on naturalistic data, as it constructs both a concrete lexicon and abstract beliefs about the correspondence between verb form and meaning.

Syntactic bootstrapping

On the syntactic bootstrapping account, children analyze the syntactic structures in which verbs appear in order to predict aspects of their meaning not well determined by the perceptual context. At a high level, this theory is a claim about the relation between two representational spaces in the mind of the learner: the space of meanings M and the space of syntactic representations S. As such, these theories must make assumptions about the structure of these spaces. As we will see, many theories regarding the syntax–semantics link presuppose the existence of core meaning predicates such as CAUSE and BECOME (Levin and Rappaport Hovav, 2011; Pinker, 1989). While such predicates have been motivated by theoretical work elsewhere in cognitive development (see e.g. Hespos and Spelke, 2004; Muentener and Carey, 2010), the continued success of the syntactic bootstrapping paradigm

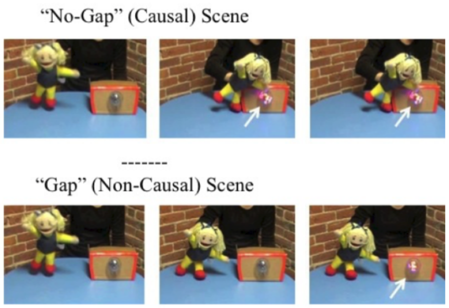

Figure 1: From Kline et al. (2017), fig. 2. Scene pairs contrast minimally in the presence or absence of a *causation* event. In the "causal" scene, the puppet moves to contact the toy, which immediately activates; in the "noncausal" scene, the puppet moves but does not contact the toy, and the toy only activates after a delay.

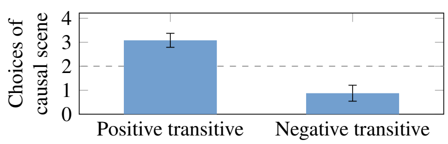

Figure 2: From Kline et al. (2017). Children point to the "causal" scene when given a positive sentence with a novel transitive verb (*can you find where she wugged the round thing?*), and to the "noncausal" scene for a negative sentence with a novel transitive verb (*... didn't wug the round thing?*).

provides orthogonal positive evidence for the reality of these structures in the mind of the child.

Corpus studies have shown how aspects of verb meaning both coarse-grained (causation and movement) and fine-grained (movement of liquids vs. movement of solids) can be predicted from the constructions in which verbs appear (Naigles and Hoff-Ginsberg, 1995; Levin, 1993). Decades of experimental evidence also support the idea that children exploit such structural relationships between a verb's syntactic behavior and its meaning. One of the most productive lines of research has focused on the correspondence between a verb's appearance in transitive syntactic constructions (*the X Ys the Z*) and a semantic predicate CAUSE (Naigles, 1990). Some of the most recent experimental evidence argues for such a fine-grained link between transitive syntax and physical causation (Kline et al., 2017). Kline et al. presented children with pairs of scenes, each involving a moving puppet and a toy which activated or lit up. An example scene pair is shown in Figure 1. While each scene pair involved similar motion events, a "causal" scene in each pair also exhibited an event of external causation, using cues known to be salient to young children (spatial and temporal continuity between an agent's action and an object's response) (Michotte, 1963; Muentener and Carey, 2010). In two-alternative forced choice test trials, children were given a sentence containing a novel verb and asked to pick the scene it referred to: either in a positive frame (*Can you find where she wugged the round thing?*) or a negative frame (*... didn't wug the round thing?*). Figure 2 shows the main effect in the experiment of Kline et al. (2017). Across several tested minimal-contrast scene pairs, children preferred to point at the causal scene when queried with the positive frame and at the non-causal scene when queried with the negative frame.

The findings of Kline et al. show that 3- and 4-year-olds latch onto a reliable relationship in English between transitive syntax and the semantic predicate CAUSE documented elsewhere in the cognitive development literature. This is a case of syntactic bootstrapping: children exploit a word's syntactic behavior in order to make guesses about its meaning.

As a broad theory regarding the construction of the lexicon, though, syntactic bootstrapping needs to eventually do quite a bit more work. Taken to its extreme, it needs to explain how each of the semantic contrasts present in a meaning space M can be explained by corresponding contrasts in a syntactic representation S. In the absence of other good accounts of verb meaning, the contrast between *chase* and *flee* and the contrast between *hit* and *kick*, for example, must be predictable from contrasts in syntactic behavior. To test the full power of syntactic bootstrapping as a theory of the construction of the lexicon, then, we must further formalize our assumptions about the structure of the syntactic space S and the meaning space M, and provide clear proof of the learnability of relations between the two spaces.

The remainder of this paper takes some first steps in that direction. We first formalize syntactic bootstrapping in a probabilistic model, showing how we can proceed from minimal assumptions about the structure of the lexicon to a model which replicates the qualitative behavior of children in syntactic bootstrapping experiments. We next show how this probabilistic model can be learned from scratch on naturalistic data, constructing both a concrete lexicon and abstract beliefs about the syntax–semantics link through only *unsupervised* experience of ambiguous language use in grounded contexts.

Related work

Most past computational models of word learning have focused on the acquisition of words with concrete referents, explaining the learning dynamics and characteristic patterns of success and failure observed in adults and children (Frank et al., 2009; Trueswell et al., 2013; Stevens et al., 2017). Our model will replicate the important structural features of these models — explicit representations of uncertainty over possible lexica, stored under strong resource limitations — and further extend to the more challenging task of acquiring verb meanings, which have either ambiguous reference or no concrete reference at all in the world of the learner.

While other computational models have been used to replicate verb learning and syntactic bootstrapping phenomena (Abend et al., 2017; Barak et al., 2014; Alishahi and Stevenson, 2008), they have been deployed only in simplified learn-

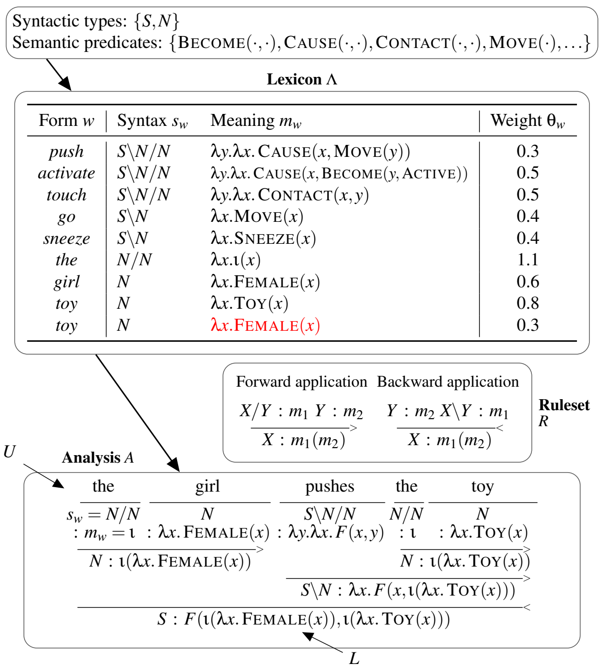

Figure 3: A CCG combines a learned lexicon Λ with a fixed ruleset R in order to yield analyses of input utterances. The bottom of the figure shows an analysis of the utterance *the girl pushes the toy* (read from top to bottom), which jointly yields syntactic and semantic representations of the sentence.

ing situations, where a learner is shown utterances explicitly paired with their ground-truth meaning representations (or a set of possible meaning representations). In contrast, our model learns in a *distantly supervised* setting: it is only explicitly told that the utterances have meanings which are *true* in the current scene, and must work out word-level meanings and utterance-level meanings on its own. Because no word-level meanings are ever explicitly presented to the learner, it must induce word meanings by searching through the infinite space of possible lambda-calculus meaning representations. This learning setting is thus qualitatively different than the direct-supervision setting studied in past bootstrapping work.

We see our model as complementary to those of Sadeghi and Scheutz (2018) and Gauthier et al. (2018), who show how more minimal syntactic representations can support specific types of early syntactic bootstrapping. Our model integrates both a full syntactic formalism and a general ability to track probabilistic links between syntactic and semantic representations. As such, the model is able to scale to the more complex syntactic bootstrapping phenomena studied in this paper, using syntactic features to resolve finer-grained features of verb meaning.

A formal model

We visualize the major details of our model in Figure 3. A learner constructs a *lexicon* Λ, associating particular word-

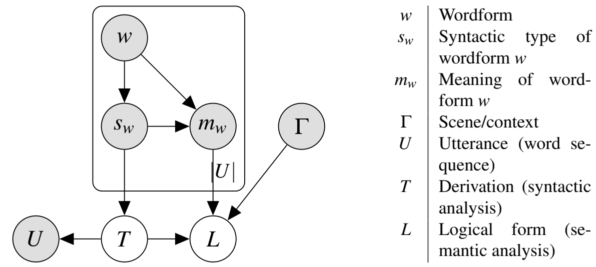

Figure 4: A generative model of an utterance U situated in a scene Γ, drawing on lexical items $(w, s_w, m_w) \in \Lambda$.

forms w with syntactic types s_w and meanings m_w.[1] The syntactic types of words are represented using the formalism of combinatory categorial grammar (CCG; Steedman and Baldridge, 2006). Word meanings are represented as expressions in a typed lambda calculus, built from core semantic predicates ranging from concrete properties (e.g. FEMALE) to abstract relations (e.g. CAUSE). These representations draw on the lexical conceptual structures often discussed in literature on the syntax–semantics link (see e.g. Levin and Rappaport Hovav, 2011).

The contents of this lexicon are combined with parsing *rules* in order to produce joint syntactic and semantic representations of full utterances. Figure 3 shows how entries from the lexicon Λ combine with a ruleset R to analyze a sentence.

The grammar's *syntactic types* describe how words combine with their arguments. These syntactic types may be of either a *primitive* type (e.g. N) or of *functional* type (e.g. S/N). Functional types combine with syntactic arguments to their left or right, eventually yielding a phrase of a particular primitive type.

The CCG rule set R, shown in the middle of Figure 3, specifies these combination rules.[2] Figure 3 (lower section) shows how the two rules in our ruleset are used to analyze the example sentence *the girl pushes the toy*. After first retrieving lexical entries for each of the tokens in the sentence (top row), we iteratively run the application rules, composing functional types with primitive types to their left or right. Whenever such syntactic composition occurs, we likewise unify the corresponding semantic expressions by function application.

Each CCG analysis yields a tree structure (bottom of Figure 3) whose root contains the syntactic type and semantic analysis of the entire input string. We call this final semantic representation the *logical form* of a sentence, and the particular tree structure of rule applications the *derivation* (analogous to a syntactic parse). We let $A = \langle L, T \rangle$ denote the full analysis of an utterance, where L is the logical form and T is the derivation.

[1] This walkthrough involves a minimal amount of equations, focusing instead on applications to concrete word learning problems. Model details are provided in the appendix of this paper.

[2] See Steedman and Baldridge (2006) for a full description.

Scene	Events
Γ_1	CAUSE(*girl*, BECOME(*toy*, *active*))
	CONTACT(*girl*, *toy*); MOVE(*girl*)
Γ_2	BECOME(*toy*, *active*); MOVE(*girl*)

Table 1: Two sample scene representations from our model of the two-alternative forced choice test trial of Kline et al. (2017).

We next design a minimal probabilistic model on top of this CCG formalism which can realize, among other things, the behavior of the children in the experiment of Kline et al. (2017). Our model adapts past work on probabilistic CCGs (see e.g. Zettlemoyer and Collins, 2007; Artzi and Zettlemoyer, 2013), adding a critical inductive bias linking syntactic and semantic representations within the lexicon. We illustrate the model as a plate diagram in Figure 4, and walk through its behavior in the following paragraphs.

We will walk through this model in the context of the Kline et al. (2017) paradigm, showing how it can realize the subjects' observed behavior. For the rest of this section, we assume the provisional lexicon shown in the top of Figure 3, associating particular words with candidate syntactic types and meanings. Later, we will remove this assumption and show how such a lexicon can be learned from experience alone.

Consider an utterance $U =$ *the girl pushes the toy* given in a grounded context Γ.[3] We can combine the CCG framework with our weighted lexicon to compute the probability of an arbitrary analysis:

$$P(A = \langle L, T \rangle \mid \Lambda, \Gamma) \propto P(\Gamma) P(L \mid \Gamma) \exp(\sum_{(w, s_w, m_w) \in T} \theta_w) \quad (1)$$

where $P(\Gamma)$ is a uniform prior over potential contexts, and $P(L \mid \Gamma)$ is one only when a logical form L is true of the context Γ. The final term in the above equation does the majority of the work, combining the weights θ_w of lexical entries involved in the derivation T. The lexicon in Figure 3 licenses multiple analyses of the sentence *the girl pushes the toy*, since it contains two candidate entries for the word *toy*. Equation (1) can be used to rank the resulting analyses — one of which is shown in the bottom section of Figure 3 — according to the constituent lexical weights θ_w.

In the experiment of Kline et al. (2017), a child hears the utterance $U =$ *the girl gorps the toy* and is asked to pick which of two scenes Γ_1, Γ_2 the utterance refers to. We represent the scenes as lists of propositions like those in Table 1.

Unlike our previous example, this utterance contains a novel word which has no corresponding entries in the lexicon. We must induce candidate syntactic types s_w and meanings m_w using the remainder of the probabilistic model.

We begin by enumerating the possible syntactic types s_w of the novel word. Given the contents of the provisional lexicon Λ (shown in the top left of Figure 3) and our parsing ruleset, there is just one syntactic analysis of *gorps* which yields a valid parse. This parse has the same structure as that shown in bottom section of Figure 3. The parse assigns the word the syntactic type $S \backslash N / N$: the syntactic type of a transitive verb.[4]

We next make predictions about the candidate meanings of *gorps*. This prediction process is visualized in Figure 5. We begin by sampling meanings m_w conditioned on the possible syntactic representations s_w. This is the point at which syntactic bootstrapping plays a critical role: the model calculates a distribution $P(m_w \mid s_w = S \backslash N / N)$, which we expect should favor meanings involving the predicate CAUSE:

$$P(\text{predicate}_i \mid s_w) \propto C + \sum_{\substack{(s_i, m_i, \theta_i) \in \Lambda \\ : s_w = s_i \wedge \text{predicate} \in m_i}} \theta_i \quad (2)$$

$$P(m_w \mid s_w) \propto \prod_{\text{predicate}_i \in m_w} P(\text{predicate}_i \mid s_w) \quad (3)$$

where C is a smoothing constant, fit as a hyperparameter. Equation (2) aggregates the total weight mass in the lexicon allocated to any particular predicate for lexical entries with syntactic type s_w. The product term of Equation (3) combines these individual predicate probabilities in order to score possible complete meanings m_w of the word *gorps*. The left panel of Figure 5 shows a ranked list of meanings computed by this equation under our provisional lexicon.

Each candidate meaning and syntactic representation of the word *gorps*, when combined with the rest of the words in the sentence, yields a full syntactic derivation T and logical form L. These utterance-level meaning representations are scored based on the scene Γ. Here we incorporate the critical constraint that logical forms L must consist of messages which are *true* of the scene Γ. This effectively filters the candidate complete meanings L, yielding a renormalized distribution over full sentence meanings as shown in the middle panel of Figure 5.

We can combine the above distributions in order to perform the critical inverse inference $P(\Gamma \mid U, \Lambda)$: which scene does the utterance *the girl gorps the toy* refer to? This distribution is computed via Bayes' rule, yielding the posterior distribution shown in the right panel of Figure 5. The positive sentence containing the novel word *gorps* is predicted to refer to the scene with a salient causation event. By a similar logic as shown in this walkthrough, the negative sentence *the girl doesn't gorp the toy* is taken to refer to the scene missing the salient causation event.

This section has demonstrated how the probabilistic model sketched in Figures 3 and 4 reproduces syntactic bootstrapping behavior, using the transitive syntax of novel words to predict meanings containing the semantic predicate of CAUSE. The model integrates the CCG parsing formalism with a statistical mechanism for tracking the relations be-

[3] Contexts will become relevant later in the paper. See Table 1 for an example context representation.

[4] In cases where there are multiple syntactic types for a novel word, they are scored according to a distribution $P(s_w \mid \Lambda)$, given in Equation (9).

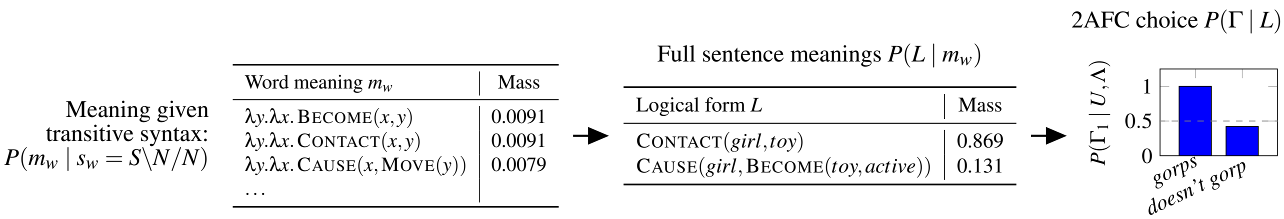

Figure 5: Computation of the meaning of a novel word *gorps* in the sentence *the girl gorps the toy* proceeds in three steps: word meanings are enumerated according to a distribution $P(m_w \mid s_w)$ (Equation (3)), full sentence meanings L are produced via CCG parsing, and the candidate scenes Γ_1, Γ_2 are scored according to which sentence meanings are true of which scene.

tween syntactic structures and their semantic correlates, helping the learner to make predictions about the meanings of novel words.

Learning

The previous section assumed that the learner already possessed a knowledge state as given in Figure 3, where wordforms like *girl* and *push* already have correct meaning representations. In this section, we show how such a lexicon can be acquired across multiple instances of ambiguous language use in context, in a manner that requires minimal long-term memory capacity and remains robust to noise in the input.

We expose our model to a sequence of observations $O = (\langle U_i, \Gamma_i \rangle)$ of utterances U_i grounded in particular scenes Γ_i. We proceed by observing each data point O_i in sequence and updating a lexicon Λ, inducing novel lexical entries as necessary and updating weights θ_w in the lexicon. The learner never directly observes the mapping between words and their referents, or between sentences and their meanings. The task of the learner is to derive word meanings, and methods for composing words, such that each utterance U_i is true in its context Γ_i.

We also constrain our word learner to encode only a limited number of lexical entries per word at all times. We label this limit ℓ, and evaluate its influence as a free parameter in the following experiments. Concretely, after each observation O_i, we retain only the ℓ highest-weight lexical entries per wordform.

Let Λ_i be a learner's lexicon representation before observing the example O_i. Suppose that the utterance U_i is observed in a context Γ_i which contains a novel word $w =$ *gorps*: *the girl* ***gorps*** *the toy*. The machinery already presented in the previous section can be used to induce candidate novel meanings for the word *gorps*. In order to support incremental learning, we include an additional *weight update* step after each utterance is observed. Given the utterance U_i, we update the weights of each lexicon entry in order to increase the probability of observing the utterance under the model given in Figure 4. Further details on the learning algorithm are given in the appendix.

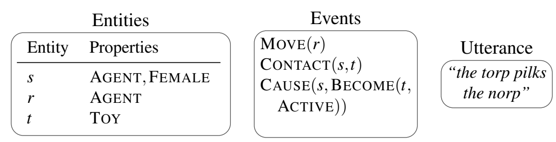

Figure 6: An example observation. Utterances refer to objects (*the norp*) or events (*the torp pilks the norp*).

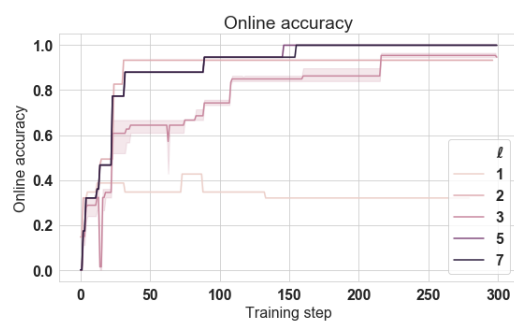

Figure 7: Online accuracy in predicting sentence meanings for word learning models with different numbers of allowed stored meanings ℓ. Shaded regions represent 95% CI.

Experiments

We deploy the above learning model on a synthetic dataset in which short utterances pick out objects, events, and relations in a simulated environment. This environment is similar to those used in artificial intelligence research on visual question answering (see e.g. Johnson et al., 2017), but contains more complex utterances which make reference to abstract events and relations (such as causation, state change, and movement). Figure 6 shows an example scene–utterance pair drawn from this dataset.

We generate observations O_i by first sampling a context Γ_i. Each context contains a random number of entities (agents and objects), and a random number of events relating those entities, structured as propositions like those shown in Table 1. Contexts always contain multiple simultaneous events, such that the learner is only ever exposed to ambiguous and indirect observations of sentence meaning.

Each entity and event is assigned a fixed random wordform throughout the experiment, and utterances are generated by combining the wordforms for the involved entities and events according to pre-designed templates. For example, if we sample a scene which contains an event $\text{CAUSE}(girl, \text{MOVE}(toy))$, we might generate an utterance *the torp pilks the norp*, where *torp* refers to female agents, *norp* refers to toys, and the whole sentence must pick out the complete event structure. We also randomly generate *negative* utterances, where verbs are modified by words to their immediate left who function to negate the overall sentence meaning. For example, *the torp doesn't pilks the norp* has the meaning $\neg\text{CAUSE}(girl, \text{MOVE}(toy))$.[5]

The task of the learner is to infer the meanings of each of these words by observation of random scenes $O_i = \langle U_i, \Gamma_i \rangle$. While we have access to the ground-truth correspondence between sentences and their full logical forms during scene generation, this mapping is not provided to the learner.

We evaluate the learner's lexicon acquisition by two metrics: 1) its accuracy in predicting the ground-truth semantic representations of test sentences, and 2) its accuracy in the syntactic bootstrapping two-alternative forced choice task of Kline et al. (2017). Figure 7 shows the model's performance on the first across learning time, as the model is incrementally exposed to more examples O_i. Both graphs contrast models with different settings of the hyperparameter ℓ, which controls the maximum number of entries that can be stored across observations for any wordform in the lexicon. For all settings of $\ell > 1$, the model reaches high performance within 100 examples. All models reach perfect performance on the second syntactic bootstrapping 2AFC task after just a few examples: the correct acquisition of just one or two transitive verbs is enough to support the induction of a productive belief about the link between verb syntax and semantics.

The results in Figure 7 demonstrate that even highly resource-constrained Bayesian learners can acquire an accurate lexicon in a data-efficient manner. These same learners quickly derive a syntactic bootstrapping capacity from their own lexicons, supporting more efficient learning in the future.

Conclusion

This paper has presented a computational word learning model which actively tracks the correspondences between the syntactic and semantic behavior of words. We demonstrated how this framework can capture experimentally observed syntactic bootstrapping phenomena, and that such phenomena can be explained as the rational behavior of a cross-situational learner exposed to a corpus of naturalistic data. Critically, both word learning (of nouns and verbs) and also the acquisition of the high-level syntactic bootstrapping behavior still go through given substantial long-term memory constraints, in which models store just a few candidate interpretations per wordform in their language.

As a computational model of acquisition, this framework makes predictions about how people should interpret and generalize novel words. Our framework allows us to make rigorous and explicit statements about the structure of the mental representational spaces underlying these generalizations. In ongoing work, we are using the same model class presented in this paper to detect candidate links between word syntax and word semantics which a rational learner can (and should) exploit.

Acknowledgments

The authors gratefully acknowledge support from the MIT Sensetime Alliance, MIT Quest for Intelligence, and the Open Philanthropy Foundation.

Model details

This final section provides mathematical details on the model for completeness.

Reading from Figure 4, the probability of a full utterance is:

$$P(U \mid \Gamma, \Lambda) \propto P(\Gamma) \sum_{L,T} P(T \mid \Lambda) P(U \mid T) P(L \mid T, \Gamma, \Lambda) \tag{4}$$

We assume a uniform prior over scenes $P(\Gamma)$, and let $P(U \mid T)$ be 1 exactly when the span of T is equivalent to U, and zero otherwise. Lastly, we define the probability of a logical form L in terms of the derivation T and context Γ as follows:

$$P(L \mid T, \Gamma, \Lambda) = P(L \mid T, \Lambda) P(L \mid \Gamma) \tag{5}$$

$$P(L \mid T, \Lambda) \propto \mathbf{1}\{L \text{ is determined by } T, \Lambda\} \tag{6}$$

$$P(L \mid \Gamma) \propto \mathbf{1}\{L \text{ is true in } \Gamma\} \tag{7}$$

Novel word induction Given a novel word w, we resort to the full Bayesian model to make predictions about its syntactic type s_w and meaning m_w.

$$P(w \to (s_w, m_w) \mid U, \Gamma) \propto$$
$$P(s_w \mid w) P(m_w \mid s_w) P(U \mid \Gamma, \Lambda \cup (w, s_w, m_w)) \tag{8}$$

The only term not yet defined is the distribution over syntactic types $P(s_w \mid w)$. This distribution is computed by simple inspection of the lexicon. The probability mass assigned to a particular syntactic category s is proportional to the total weight assigned to entries in Λ with category s:

$$P(s_w \mid \Lambda) \propto C + \sum_{(w, s_w, m_w, \theta_w) \in \Lambda} \theta_w \tag{9}$$

where C is a smoothing constant.

As shown in the earlier model walkthrough, Equation (8) is used to initialize the weights for the lexical entries of novel words.

Weight updates Let g be the highest probability correct analysis of a sentence $\langle L, T \rangle$, and let B be the set of the k most probable incorrect analyses.[6] For each lexical entry $x_i = (w_i, s_i, m_i, \theta_i)$ with weight θ_i, we perform the following perceptron update:

$$\theta_i \mathrel{+}= \eta \left(\mathbf{1}\{x_i \in g\} - \frac{1}{|B|} \sum_{b \in B} \mathbf{1}\{x_i \in b\} \right) \tag{10}$$

where η denotes a learning rate, and $x_i \in A$ is true iff the lexical item x_i participates in the analysis A. Note that this update will only affect lexical entries with wordforms used in the utterance U_i.

[5] The training corpus is generated from a collection of 3 unique referents and 5 unique event types, each of which has one fixed referring expression. This yields a total of 51 unique utterances.

[6] Here a "correct" analysis is one which has nonzero probability under Equation (1). Note that, consistent with the cross-situational paradigm, only analyses with meanings that are true of Γ_i have nonzero probability.

References

Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., and Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*.

Alishahi, A. and Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive science*, 32(5):789–834.

Artzi, Y. and Zettlemoyer, L. (2013). Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association for Computational Linguistics*, 1:49–62.

Barak, L., Fazly, A., and Stevenson, S. (2014). Learning verb classes in an incremental model. In *CMCL*, pages 37–45.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development.

Fisher, C., Gertner, Y., Scott, R. M., and Yuan, S. (2010). Syntactic bootstrapping.

Frank, M. C., Goodman, N. D., and Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, 20(5):578–585.

Gauthier, J., Levy, R., and Tenenbaum, J. B. (2018). Word learning and the acquisition of syntactic–semantic overhypotheses. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, Madison, Wisconsin.

Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1):23–64.

Hespos, S. J. and Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, 430(6998):453.

Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Lawrence Zitnick, C., and Girshick, R. (2017). Clevr. In *CVPR*.

Kline, M., Snedeker, J., and Schulz, L. (2017). Linking language and events. *Language Learning and Development*, 13(1):1–23.

Landau, B. and Gleitman, L. R. (1985). Language and experience: Evidence from the blind child.

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

Levin, B. and Rappaport Hovav, M. (2011). Lexical conceptual structure. *Semantics*.

Michotte, A. (1963). *The perception of causality*. Basic Books.

Muentener, P. and Carey, S. (2010). Infants causal representations of state change events. *Cognitive psychology*, 61(2):63–86.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, 17(2):357–374.

Naigles, L. R. and Hoff-Ginsberg, E. (1995). Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, 31(5):827.

Pinker, S. (1989). Learnability and cognition: The acquisition of argument structure.

Sadeghi, S. and Scheutz, M. (2018). Early syntactic bootstrapping in an incremental memory-limited word learner. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Smith, L. and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*.

Steedman, M. and Baldridge, J. (2006). Combinatory categorial grammar.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., and Yang, C. (2017). The pursuit of word meanings. *Cognitive science*, 41:638–676.

Trueswell, J. C., Medina, T. N., Hafri, A., and Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, 66(1):126–156.

Zettlemoyer, L. and Collins, M. (2007). Online learning of relaxed ccg grammars for parsing to logical form. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.