Starting from Scratch in Semantic Role Labeling: Early Indirect Supervision

Michael Connor, Cynthia Fisher and Dan Roth

Abstract A fundamental step in sentence comprehension involves assigning semantic roles to sentence constituents. To accomplish this, the listener must parse the sentence, find constituents that are candidate arguments, and assign semantic roles to those constituents. Where do children learning their first languages begin in solving this problem? To experiment with different representations that children may use to begin understanding language, we have built a computational model for this early point in language acquisition. This system, Latent BabySRL, learns from transcriptions of natural child-directed speech and makes use of psycholinguistically plausible background knowledge and realistically noisy semantic feedback to improve both an intermediate syntactic representation and its final semantic role classification. Using this system we show that it is possible for a simple learner in a plausible (noisy) setup to begin comprehending the meanings of simple sentences, when initialized with a small amount of concrete noun knowledge and some simple syntax-semantics mapping biases, before acquiring any specific verb knowledge.

1 Introduction

When learning their first language, children must cope with enormous ambiguity in both the meaning and structure of input sentences. Ultimately, children must select candidate meanings by observing the world and align them with the sentence presented in the input. They must do so without already knowing which parts of the sentence refer to which parts of their conceptual representations of world events.

Michael Connor
Department of Computer Science, University of Illinois e-mail: connor2@illinois.edu

Cynthia Fisher
Department of Psychology, University of Illinois e-mail: clfishe@illinois.edu

Dan Roth
Department of Computer Science, University of Illinois e-mail: danr@illinois.edu

Even worse, the child must also identify the ways in which structural aspects of sentences, which are not clearly displayed in the surface form of the utterance, convey aspects of the relational meanings of those sentences. For example, phrase order or case marking identify the roles that particular constituents play in the sentence's meaning, thus conveying who does what to whom. Despite both of these sources of ambiguity, semantic and syntactic, children do learn to interpret sentences, and do so without detailed feedback about whether their interpretations, or their hypothesized syntactic structures, were correct. When faced with an ambiguous world, and with word-strings rather than sentence structures, how can learners begin to identify and interpret the syntactic structures of sentences?

The ambiguity of word-strings as evidence for syntax is a nearly universally recognized problem for language acquisition. But the ambiguity of scenes as evidence for sentence meaning is sometimes overlooked. To illustrate, take the sentence "The girl tickled the boy," accompanied by a scene in which a boy and girl play together, and at some point the girl does tickle the boy. Any scene offers up a host of candidate interpretations, both related and unrelated to the target event described by the sentence. These might include the boy and girl playing, the boy squirming and giggling, the girl giggling, background facts about the boy or girl that might be of interest (e.g., "You know that girl from preschool."), and so forth. Among the available construals might be some that are very difficult to tease apart based on even an extended sequence of suitable scenes. For example, scenes of 'giving' nearly always also involve 'getting', scenes of 'chasing' involve 'fleeing,' and scenes of 'putting' an object in a location also imply that the object 'goes' into that location. The basic predicate-argument semantics of a sentence are not simple descriptions of scenes, but rather express the speaker's selected perspective on that scene (e.g., Clark (1990)). It is up to the speaker to direct the attention of the child listener to the correct interpretation, through various means such as looking, gestures (Nappa et al., 2009) and *the sentence itself*.

In this chapter we develop a computational language learner that must cope with this ambiguity of both scene and sentence in learning to classify abstract semantic roles for verbal predicate arguments. This computational learner, our 'Latent BabySRL', learns from child directed speech transcripts and ambiguous semantic feedback, treating an intermediate syntactic representation as a latent structure that must be learned along with the semantic predictions. This system allows us to test various plausible sources of knowledge and representation for the child learner, showing that simple structural cues regarding the identification of nouns are necessary for disambiguating noisy semantics.

1.1 Addressing the ambiguity of sentences and scenes: Semantic and syntactic bootstrapping

A vivid illustration of the ambiguity of scenes comes from 'human simulation' experiments devised by Gleitman and colleagues to investigate word learning based on observations of accompanying scenes (Gillette et al., 1999; Snedeker and Gleitman, 2004). In these experiments, adult observers watched video-clips of mothers

interacting with toddlers. In each video-clip, the mother had uttered a common noun or verb; the soundtracks of the videos were removed and participants heard only a 'beep' at the point in each video when the target word had been spoken. The observer's task was to guess what word the mother had said. Observers saw a series of such clips for each target word; thus they had opportunities for cross-situational observation. Participants were much more accurate in guessing the target nouns than the verbs. Performance with verbs improved considerably, however, when participants also received information about the sentence structures in which the verbs occurred. These results suggest that scene observations are systematically less informative for learning verbs than for learning nouns. The referents of many concrete nouns can be identified via scene observation alone, but verb referents are typically more abstract (i.e. less 'imageable') (Gillette et al., 1999), and therefore naturally harder to observe in scenes. Efficient verb learning depends on support from sentence-structure cues.

On the other hand, despite the ambiguity of scenes, it is clear that a substantial part of the evidence required to learn a language must come from observing events. Only by observing words used in appropriate referential contexts (e.g., 'feed' when feeding is relevant, 'cookie' when cookies are relevant) could children attach appropriate semantic content to those words. For this reason, theories of language acquisition routinely assume that learning to use words and syntax in sentence interpretation is a partly supervised task, where the supervision comes from observation of world events: For each input sentence, learners use their existing knowledge of words and sentence structure to generate a possible meaning; the fit of this meaning with the referential context provides feedback for improving the child's lexical and grammatical knowledge. We would argue, however, that most theories or models of language acquisition finesse the true ambiguity of observation of world events, by assuming that the child has access to the correct interpretation of input sentences some substantial proportion of the time – often enough to support robust acquisition (e.g.,Chang et al. (2006); Pinker (1984); Tomasello (2003)).

The semantic bootstrapping theory is a special case of such accounts (Pinker, 1984, 1989). The semantic bootstrapping theory focuses on the *ambiguity of word-strings* as evidence for syntactic structure, and proposes that learners are equipped with innate links between semantic and syntactic categories and structures; these links allow them to use semantic evidence to identify words and structures that are of particular syntactic types in their native language. To take a simple example, children might infer that words referring to entities in the world are nouns, or that a phrase referring to an agent of action in the main event described by an input sentence is the grammatical subject. On this account, access to word and sentence meaning (derived from scene observations) plays a privileged role in identifying syntactic structure, with the aid of innate links between syntax and semantics. To return to our 'tickle' example, the child would use previously acquired knowledge of the content words in this sentence ('girl', 'boy', and 'tickle') to choose the relevant construal of the scene. Via semantic bootstrapping, the child would then infer that the noun-phrase naming the agent of tickling should be the grammatical subject of the sentence. The sentence "The girl tickled the boy" would then yield a data point

that the child could begin to use to determine where to find the grammatical subject in English sentences.

The syntactic bootstrapping theory (Landau and Gleitman, 1985; Naigles, 1990), in contrast, focuses on the *ambiguity of scenes*, particularly with regard to learning the abstract relational meanings of verbs and of sentence-structural devices such as word order or case marking (Gillette et al., 1999). Syntactic bootstrapping proposes that children use partial knowledge of sentence structure to select likely meanings of input sentences; by doing so they gain access to syntactic support for verb learning. Like semantic bootstrapping, syntactic bootstrapping requires that the learner have access to links between syntax and semantics; for syntactic bootstrapping to play a role in the initial creation of a lexicon and grammar, some of these links must be innate. The nature of these links is typically assumed to follow from the fundamental nature of the relational meanings of verbs (Gillette et al., 1999; Landau and Gleitman, 1985; Naigles, 1990; Fisher et al., 1989): Verbs are argument-taking predicates, and the number of semantic arguments required to play out the meaning of each verb is systematically related to the phrasal structure of sentences containing that verb (e.g., Levin and Rappaport-Hovav (2005); Pinker (1989)). In our 'tickle' example, the presence of two noun-phrase arguments in the target sentence "The girl tickled the boy" is clearly no accident, but reflects the underlying predicate-argument structure of the verb.

1.2 How could syntactic bootstrapping begin?

But given the dual problem we started with, the rampant ambiguity of both word-strings and scenes, how could any aspects of sentence structure begin to guide sentence interpretation without considerable prior learning about the syntax and morphology of the native language? The 'structure-mapping' account of the origins of syntactic bootstrapping (Fisher et al., 2010) proposes one way in which sentence structures might first guide sentence interpretation, even before children learn much about the syntax of the native language.

First, the structure-mapping account proposes that children are predisposed to align each noun in a sentence with a core semantic argument of a predicate. Given this bias, the number of nouns in the sentence becomes intrinsically meaningful to toddlers. In our 'tickle' illustration, simply identifying the target sentence as containing two nouns should prompt children to select an interpretation with two core participant roles. This simple constraint allows a skeletal representation of sentence structure, grounded in the learning of some nouns, to guide sentence interpretation essentially from the start – and to do so without requiring prior knowledge of verb meanings. This simple inference would yield a probabilistic distinction between transitive and intransitive sentences, increasing the probability that children interpret an input sentence as its speaker intended, despite the ambiguity of scenes. In turn, this increased accuracy in sentence interpretation puts the child in a better position to obtain useful information from the observed scene about other aspects of the meaning of the sentence. Such experiences provide useful information about 'tickle,' and about the interpretation of English sentences more generally.

Second, the structure-mapping account, like any form of syntactic bootstrapping, assumes that children represent their experience with language in usefully abstract terms. These abstract representations both give children access to the proposed innate bias to align nouns with participant-roles (Yuan et al., ress), and permit rapid generalization of language-specific learning to new sentences and new verbs (Gertner et al., 2006; Pinker, 1984). As a result, each advance in learning the syntactic choices of the native language offers new constraints on verb and sentence interpretation. The structure-mapping account proposes that even skeletal representations of sentence structure grounded in a set of nouns provide a preliminary format for further learning about the syntax of the native language (see also (Bever, 1970)). To illustrate, experiences like the one sketched in our 'tickle' example, given abstract representations of both (partial) sentence structure and semantic roles, could provide the learner with evidence that the first of two noun arguments is an agent of action, and the second is a patient or recipient of action.

This process exemplifies the kind of iterative, opportunistic learning from partial knowledge that inspired the term 'bootstrapping'. It naturally incorporates aspects of both semantic and syntactic bootstrapping (Gillette et al., 1999): Children are assumed to identify the referents of some concrete nouns via a word-to-world mapping unaided by syntactic bootstrapping. As a result, early vocabularies tend to be dominated by nouns (Gentner, 2006). Children then assume, by virtue of the referential meanings of these nouns, that the nouns are candidate arguments of verbs. This is a simple form of semantic bootstrapping, requiring the use of built-in assumptions about syntax-semantics links to identify the grammatical function of known words – nouns in particular (Pinker, 1984). In this way, an initial noun vocabulary grounds a preliminary estimate of the syntax of the sentence, which in turn permits further word and syntax learning, via syntactic bootstrapping.

In this chapter we use a Semantic Role Labeling (SRL) task (Carreras and Màrquez, 2004) based on child-directed speech (CDS) to model these initial steps in syntactic bootstrapping. Computational models of semantic role labeling face a learning problem similar to the one children face in early sentence comprehension: The system learns to identify, for each verb in a sentence, all constituents that fill a semantic role, and to determine their roles, such as agent, patient or goal. Our 'BabySRL' system (Connor et al., 2008, 2009, 2010) learns to predict the semantic roles of verbs' arguments in input sentences by directly implementing the assumptions of the 'structure-mapping' account. That is, the model 1) assumes that each noun is a candidate argument of a verb and 2) models the semantic prediction task with abstract role labels and abstract (though partial) sentence-representations grounded in a set of nouns. Our goals in implementing these assumptions in a computational model of semantic role labeling were to test the main claims of our account by explicitly modeling learning based on the proposed skeletal description of sentence structure, given natural corpora of child-directed speech. We equipped the model with an unlearned bias to map each noun onto an abstract semantic role, and asked whether partial representations grounded in a set of nouns are useful as a starting point in learning to interpret sentences. We used English word-order as a first case study: Can the BabySRL learn useful facts about English sentence-

interpretation, such as that the first of two nouns tends to be an agent? Crucially, in the present modeling experiments we asked whether learning that begins with the proposed representational assumptions can be used to improve the skeletal sentence representations with which the learner began.

In carrying out the simulations described here, our main preoccupation has been to find ways for our model to reflect both the ambiguity of scene-derived feedback about the meaning of input sentences and the ambiguity of word-strings as evidence for syntactic structure. Like the major theoretical accounts of language acquisition briefly discussed above, computational language-learning systems (including both those in the the Natural Language Processing (NLP) tradition and more explicitly psycholinguistically-inspired models) often rely on implausibly veridical feedback to learn, both in divining syntactic structure from a sentence and in fitting a meaning to it. For example, the state-of-the-art SRL system which we used as a baseline for designing our BabySRL (Punyakanok et al., 2008), like other similar systems, models semantic-role labeling in a pipeline model, involving first training a syntactic parser, then training a classifier that learns to identify constituents that are candidate arguments based both on the output of the preceding syntactic parser and on direct feedback regarding the identity of syntactic arguments and predicates. Features derived from the output of this closely supervised syntactic predicate-argument classifier then serve as input to a separate semantic-role classifier that learns to assign semantic roles to arguments relative to each predicate, given feedback about the accuracy of the role assignments. At each level in this traditional pipeline architecture, the structure that is learned is not tailored for the final semantic task of predicting semantic roles, and the learning depends on the provision of detailed feedback about both syntax and semantics. In essence, whereas children learn through applying partial knowledge at multiple levels of the complex learning and inference problem, successful computational learners typically require incorporating detailed feedback at every step. Therefore our first steps in developing the BabySRL have been to simplify the representations and the feedback available at each step, constrained by what we argue is available to children at early points in language learning (see below).

In the present work we built a computational system that treats a simple form of syntax as a hidden structure that must be learned jointly with semantic role classification. Both types of learning are based on the representational assumptions of the structure-mapping account, and on the provision of high-level, but varyingly ambiguous, semantic feedback. To better match the learning and memory capabilities of a child learner, we implemented our learning in an online, sentence-by-sentence fashion.

With this system we aim to show that:

- Nouns are relatively easy to identify in the input, using distributional clustering and minimal supervision.
- Once some nouns are identified as such, those nouns can be used to identify verbs based on the verbs' argument-taking behavior.
- The identification of nouns and verbs yields a simple linear sentence structure that allows semantic-role predictions.

(a) Sentence	The girl tickled the boy .	
(b) Semantic Feedback	A0 A1	
(c) Syntactic Structure	N V N	
(d) Feature Representation	*girl* argument:girl predicate:tickled NPat: 1st of 2 Ns VPos:Before Verb	*boy* argument:boy predicate:tickled NPat: 2nd of 2 Ns VPos: After Verb

Table 1 Example input and feedback representation for the original BabySRL system. For each training sentence (a), gold standard semantic feedback (b) provided true abstract role labels for each argument, and gold standard part-of-speech tagging provided true identification of the nouns and verbs in the sentence (c). Each noun was treated as an argument by the semantic-role classifier; in the input to this classifier, nouns were represented (features (d)) by the target argument and predicate themselves, and features indicating the position of each noun in a linear sequence of nouns (NPatttern or NPat, e.g., 1st of 2 nouns, 2nd of 2 nouns) and its position relative to the verb (VPosition or VPos). Section 4.1.1 will further describe these features.

- The skeletal sentence structure created via minimally supervised noun identification provides constraints on possible sentence structures, permitting the Latent BabySRL to begin learning from highly ambiguous semantic-role feedback.

2 BabySRL and Related Computational Models

In our previous computational experiments with the BabySRL, we showed that it is possible to learn to assign abstract semantic roles based on shallow sentence representations that depend only on knowing the number and order of nouns; the position of the verb, once identified, added further information (Connor et al., 2008). Table 1 gives an example of the representations and feedback that were originally used to drive learning in the BabySRL. In our first simulations (Connor et al., 2008), full (gold standard) semantic-role feedback was provided along with a shallow syntactic input representation in which nouns and verbs were accurately identified. This skeletal representation sufficed to train a simple semantic role classifer, given samples of child-directed speech. For example, the BabySRL succeeded in interpreting transitive sentences with untrained (invented) verbs, assigning an agent's role to the first noun and a patient's role to the second noun in test sentences such as "Adam krads Mommy". These first simulations showed that representations of sentence structure as simple as 'the first of two nouns' are useful as a starting point for sentence understanding, amid the variability of natural corpora of child-directed speech.

However, the representations shown in Table 1 do not do justice to the two sources of ambiguity that face the human learner, as discussed above. The original BabySRL modeled a learner that already (somehow) knew which words were nouns and in some versions which were verbs, and also could routinely glean the true interpretation of input sentences from assumed observation of world events. These are the kinds of input representations that make syntactic and semantic bootstrapping unnecessary (in the model), and that we have argued are not available to the novice

learner. Therefore in subsequent work, we began to weaken these assumptions, reducing the amount of previous knowledge assumed by the input representations and by the semantic-role feedback provided to the BabySRL. These next steps showed that the proposed simple structural representations were robust to drastic reductions in the integrity of the semantic-role feedback (when gold-standard semantic role feedback was replaced with a simple animacy heuristic for identifying likely agents and non-agents; (Connor et al., 2009)) or of the system for argument and predicate identification (when gold standard part-of-speech tagging was replaced with a minimimally-supervised distributional clustering procedure; (Connor et al., 2010)). In this chapter we develop a system that learns the same semantic role labeling task when given input representations and feedback that in our view more closely approximate the real state of the human learner: semantic feedback that is dramatically more ambiguous, coupled with the need to infer a hidden syntactic structure for sentences presented as word-sequences, based on the combination of bottom-up distributional learning with indirect and ambiguous semantic feedback.

Much previous computational work has grappled with core questions about the earliest steps of language acquisition, and about links between verb syntax and meaning. Here we briefly review some major themes in these literatures, focusing on the range of assumptions made by various classes of models. In particular, we specify what problems of language learning each class of models attempts to solve, and what input and feedback assumptions they rely on to do so. As we shall see, the field has largely kept separate the learning of syntactic categories and structures on the one hand, and the learning of syntax-semantics links on the other. Few models attempt to combine the solutions to both of these problems, and we would argue that none simultaneously reflect the two central ambiguity problems (of sentence and scene input) that face the learner.

First, a large and varied class of computational models explores the use of distributional learning in a constrained architecture to permit the unsupervised identification of syntactic categories or structures. For example, clustering words based on similar distributional contexts (e.g., preceding and/or following words) results in word-classes that strongly resemble syntactic categories (e.g., Brown et al. (1992); Elman (1991); Johnson (2007); Mintz et al. (2002); Mintz (2003)). In these systems, the text itself is typically the only input to the learner, but the nature of the classes also depends on the model's assumptions about how much context is available, and how (and how many) clusters are formed. Several influential recent models have extended such distributional analysis techniques to discover the constituent structure of sentences, and hierarchical dependencies between words or constituents (e.g. Bod (2009); Klein and Manning (2004); Solan et al. (2005); Waterfall et al. (2010)). These models again are unsupervised in the sense that they receive only word-sequences (or word-class sequences) as input, with no direct feedback about the accuracy of the structures they infer. They are also constrained by various assumptions about the nature of the structures to be uncovered (e.g., binary hierarchical structures), and by pressures toward generalization (e.g., minimum description length assumptions). The constraints imposed constitute the model's fragment of Universal Grammar. These models inherit a long-standing focus on the importance of dis-

tributional analysis in linguistics (Harris, 1951; Yang, 2011); jointly, such models demonstrate that appropriately constrained distributional analysis yields powerful cues to grammatical categories and structures. However, these models create *unlabeled* categories and structures, yielding no clear way to link their outputs into a grammar, or a system for interpreting or producing sentences. For the most part, distributional learning models have not been linked with models of sentence processing (though we will discuss one exception to this rule below). This is one of the goals of the current work, to link bottom-up distributional learning with a system for semantic-role labeling.

Second, a distinct class of models tackles the learning of relationships between syntax and semantics. A prominent recent approach is to use hierarchical Bayesian models to learn flexible, multi-level links between syntax and semantics, including syntactic-semantic classes of verbs and abstract verb constructions (e.g., Parisien and Stevenson (2010); Perfors et al. (2010)), the abstract semantic roles that are linked with particular argument positions within verb frames or constructions (Alishahi and Stevenson, 2010), and verbs' selection restrictions (Alishahi and Stevenson, 2012). These models address fascinating questions about the nature and representation of links between form and meaning. However, they do not attempt to address the ambiguity of either the sentence or scene input for the novice learner. Models in this class typically begin with input sentences that are already specified in both syntactic and semantic terms.

For example, Table 2 presents an input representation for the sentence "Sarah ate lunch" as presented to the models of Alishahi and Stevenson (2010, 2012). The syntactic part of this representation includes the identity of the verb, and the identity, number, and order of the verb's arguments. The semantic part is constructed based on hand-annotated verb usages and semantic properties extracted from the WordNet hierarchy (Miller et al., 1990). The semantic representations provide both lexical-semantic features of the arguments and verb (e.g., that 'Sarah' is female) and features representing the role each argument plays in the event denoted by the verb (e.g., Sarah's role is volitional); the role features are derived from theoretical descriptions of the semantic primitives underlying abstract thematic roles (e.g.,Dowty (1991)). Thus, like the original BabySRL described above, models in this class represent a learner that has already acquired the grammatical categories and meanings of the words in the sentence, and can identify the relational meaning of the sentence. In essence, these models assume that identifying basic aspects of the syntactic structure of the sentence, and identifying the sentence's meaning, are separate problems that can be addressed as precursors to discovering links between syntax and semantics. The key argument of both the syntactic and semantic bootstrapping theories is that this is not true; on the contrary, links between syntax and semantics play a crucial role in allowing the learner to identify the syntax of the sentence, its meaning, or both (Pinker, 1984; Landau and Gleitman, 1985).

An influential model by Chang and colleagues is an exception to the rule that distributional-learning models are kept separate from higher-level language processing tasks. Chang et al. (2006) implemented a model that learns to link syntax and semantics without predefined syntactic representations. Chang et al. modeled learn-

(a) Sentence	Sarah ate lunch .
(b) Syntactic Pattern	arg1 verb arg2
(c) Semantic Properties	verb: {act, consume} eat arg1 lexical: {woman, adult female, female, person ...} role: {volitional, affecting, animate ...} arg2 lexical: {meal, repast, nourishment ...} role: {non-independently exist, affected ...}

Table 2 Example input sentence and extracted verb frame from Alishahi and Stevenson (2010). The model learns to identify the subset of lexical and role features that are characteristic of each argument position within similar verb usages, thus learning abstractions such as 'agent' and 'patient'. The model assumes knowledge of the meanings of individual verbs and their arguments (using the WordNet hierarchy and hand-constructed event-role representations), and also syntactic knowledge of the identity of the verb and arguments, and the number and order of arguments in the sentence.

ing in a system that yokes a syntactic sequencing system consisting of a simple recurrent network (SRN), to a distinct message system that represents the meaning of each input sentence. The message system represents each sentence's meaning via lexical-semantic representations that specify what particular actions and entities are involved in the meaning, bound to abstract event-role slots (action, agent, theme, goal ...) that specify how many and what argument-roles are involved. In a typical training trial, the model is presented with a fixed message for the sentence, and a sequence of words conveying that message. The model tries to predict each next word in the sentence from the previous words, based on prior learning in the SRN and knowledge of the message. A key feature of this model is that the hidden units of the SRN are linked by learnable weights to the abstract event-role slots of the message system, but not to the lexical-semantic part of the message. This "Dual-Path" architecture keeps lexical-semantic information out of the syntactic sequencing system, thus ensuring that the model formulates abstract rather than word-specific syntactic representations in its hidden units. This system is unique in that it models online sentence processing, making predictions that change word by word as the sentence unfolds; thus, unlike the other models discussed in this section (including the BabySRL), it can be used to investigate how syntactic learning depends on the order in which information becomes available in sentences. The current effort shares with the dual-path model the linking of distributional learning into a system that learns to link syntax and semantics. However, the dual-path model creates syntactic representations by assuming the child already has accurate semantic representations of the input sentences. This model therefore resembles semantic bootstrapping in its reliance on meaning to drive syntax learning in a constrained architecture. We sought to create a model in which the problems of sentence and scene ambiguity could be solved jointly, allowing very partial syntactic constraints to help select a meaning from an ambiguous scene.

In jointly addressing these two types of ambiguity, our work could be viewed as analogous to a recent model of the task of word segmentation, which logically precedes the sentence-interpretation task we examine here. Johnson et al. (2010) present a computational model that jointly learns word segmentation along with word-referent mappings; they demonstrate synergistic benefits from learning to solve these problems jointly. Here we try to apply a similar insight at a different level of analysis, to learn about the structure of the sentence (identifying arguments and predicates) along with a semantic analysis of the sentence (identifying semantic roles). These high-level processing steps of course also depend on word-segmentation success; although we do not yet incorporate this step, it could be argued that additional benefits could be achieved by learning jointly across all levels of language processing, from word segmentation through sentence-structure identification to semantic interpretation.

In learning semantic role labeling, it is well known that the parsing step which gives structure to the sentence is pivotal to final role labeling performance (Gildea and Palmer, 2002; Punyakanok et al., 2008). Given the dependence of semantic role labeling on parsing accuracy, there is considerable interest in trying to learn syntax and semantics jointly, with two recent CoNLL shared tasks devoted to this problem (Surdeanu et al., 2008; Hajič et al., 2009). In both cases, the best systems learned syntax and semantics separately, then applied them together, so at this level of language learning the promise of joint synergies has yet to be realized.

3 Model of Language Acquisition

As noted earlier, Semantic Role Labeling is an NLP task involving identifying and classifying the verbal predicate-argument structures in a sentence, assigning semantic roles to arguments of verbs. Combined with the development of robust syntactic parsers, this level of semantic analysis should aid other tasks requiring intelligent handling of natural language sentences, including information extraction and language understanding. A large literature exploring the SRL task began to emerge with the development of the PropBank semantic annotated corpora (Kingsbury and Palmer, 2002; Palmer et al., 2005) and the introduction of the CoNLL (Annual Conference on Computational Natural Language Learning) shared task SRL competitions (Carreras and Màrquez, 2004, 2005). For a good review of the SRL task along with a summary of the state of the art, see Màrquez et al. (2008).

To illustrate, (1) is a sentence from PropBank:

(1) Mr. Monsky *sees* much bigger changes ahead.

The SRL task is to identify the arguments of the verb "sees" and classify their roles in this structure, producing the labeling in (2). In example (2), square brackets mark the identified arguments; A0 (sometimes written as Arg-0) represents the agent, in this case the seer, A1 (also Arg-1) represents the patient, or that which is seen, and AM-LOC is an adjunct that specifies the location of the thing being seen.

(2) [$_{A0}$ Mr. Monsky] *sees* [$_{A1}$ much bigger changes] [$_{AM-LOC}$ ahead] .

PropBank defines two types of argument roles: core roles A0 through A5, and adjunct-like roles such as the AM-LOC above[1]. The core roles in the PropBank coding scheme represent a strong assumption about the nature of semantic roles (Palmer et al., 2005); this assumption is also a key assumption of the structure-mapping account. That is, the core role labels (especially A0 and A1) are assumed to be abstract semantic roles that are shared across verbs, although the precise event-dependent meanings of the roles depends on the verb. For example, the argument of each verb whose role is closest to a prototypical agent (Dowty, 1991) is marked as A0; this would include the seer for 'see', the giver for 'give', and so forth. The argument whose role is closest to a prototypical patient is designated A1; this includes the thing seen for 'see', the thing given for 'give', and so forth. These role assignments are given for each verb sense in the frame files of PropBank. Each frame file has a different frame set for each sense of a verb that specifies and defines both the possible roles and the allowable syntactic frames for this verb sense. The across-verb similarity of roles sharing the same role-label is less obvious for the higher-numbered roles. For example, A2 is a source for 'accept', and an instrument for 'kick'.

3.1 CHILDES Training Data

One goal of the BabySRL project was to assess the usefulness of a proposed set of initial syntactic representations given natural corpora of child directed speech. Therefore we used as input samples of parental speech to three children (Adam, Eve, and Sarah; (Brown, 1973)), available via CHILDES (MacWhinney, 2000). The semantic-role-annotated corpus used in this project consists of parental utterances from sections Adam 01-23 (child age 2;3 - 3;2), Eve 01-20 (1;6 - 2;3), and Sarah 01-90 (2;3 - 4;1). All verb-containing utterances without symbols indicating disfluencies were automatically parsed with the Charniak parser (Charniak, 1997) and annotated using an existing SRL system (Punyakanok et al., 2008); errors were then hand-corrected. The final annotated sample contains 15,148 sentences, 16,730 propositions, with 32,205 arguments: 3951 propositions and 8107 arguments in the Adam corpus, 4209 propositions and 8499 arguments in Eve, and 8570 propositions and 15,599 arguments in Sarah.

3.1.1 Preprocessing and Annotation

During preprocessing of the CDS transcripts, only utterances from the Mother and Father were used. Other adults were typically present, including the researchers who

[1] In our corpus the full set of role labels is: A0, A1, A2, A3, A4, AM-ADV, AM-CAU, AM-DIR, AM-DIS, AM-EXT, AM-LOC, AM-MNR, AM-MOD, AM-NEG, AM-PNC, AM-PRD, AM-PRP, AM-RCL, AM-TMP

collected the data, but we focused on parental speech because we considered it most likely to be typical CDS. Because our goal was to create a corpus for studying input for language learning, we made no attempt to annotate the children's speech.

In the process of annotation, as noted above we removed all parental utterances that contained symbols indicating unintelligible speech, or that did not contain a verb. In addition, after pilot annotation of utterances to one child (Eve), additional guidelines were set, especially in regard to what constituted a main or auxiliary verb. In particular, we decided not to annotate the verb 'to be' even when it was the main verb in the sentence. As a result of these decisions, although there were 45,166 parental utterances in the sections annotated, only 15,148 were parsed and annotated, fewer than 34% of all utterances. This may seem like a surprisingly small proportion of the input to the children, but many of the ignored utterances were single-word exclamations ("Yes", "What?", "Alright," etc.), or were phrasal fragments that did not contain a main verb ("No graham crackers today." "Macaroni for supper?"). Such fragments are common in casual speech, and particularly so in speech to children. For example, in another corpus of child-directed English, only 52% of the utterances were full clauses (the rest were phrasal fragments or single-word exclamations), and a substantial proportion of the full clauses had 'to be' as their main verb, as in "Who's so tall?" (Fisher and Tokura, 1996).

Annotators were instructed to follow the PropBank guidelines (Palmer et al., 2005) in their semantic annotations, basing decisions on PropBank's previously-identified verb frames. If no frame existed for a specific verb (such as "tickle", found in CDS but not in the newswire text on which PropBank was developed), or a frame had to be modified to accommodate uses specific to casual speech, then the annotators were free to make a new decision and note this addition[2].

In the main experiments reported in this chapter we used samples of parental speech to one child (Adam; (Brown, 1973)) as training and test data, sections 01-20 (child age 2;3 - 3;1) for training, and sections 21-23 for test. To simplify evaluation, we restricted training and testing to the subset of sentences with a single predicate (over 85% of the annotated sentences). Additionally, in argument identification we focus on noun arguments, as will be described below. This omits some arguments that are not nouns (e.g., 'blue' in "Paint it blue."), and some semantic roles that are not typically carried by nouns. The final annotated sample contained 2882 sentences, with 4778 noun arguments.

3.2 Learning Model

The original architecture for our BabySRL was based on the standard pipeline architecture of a full SRL system (Punyakanok et al., 2008), illustrated in the top row of Figure 1. The stages are: (1) Parsing of the sentence, (2) Identifying potential arguments and predicates based on the parse, (3) Classifying role-labels for each

[2] Corpus, decision files and additional annotation information available at `http://cogcomp.cs.illinois.edu/~connor2/babySRL/`

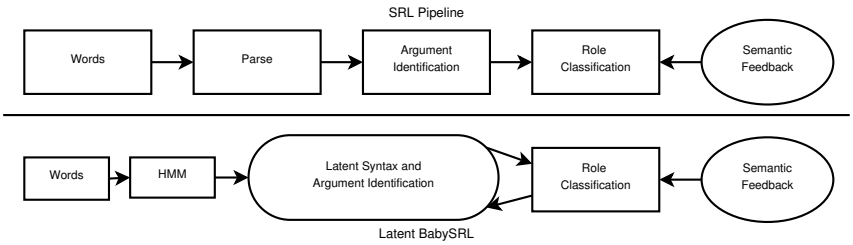


Fig. 1 Comparison of basic architecture of traditional pipeline approach for Semantic Role Labeling versus Latent BabySRL approach introduced here.

potential argument, trained using role-labeled text. Each stage depended on the accuracy of the previous stages: argument identification depends on a correct parse, role labeling depends on correct arguments.

The key intuition of the Latent BabySRL described here is that we can use the task of semantic role labeling to generate and improve the intermediate syntactic representations that support this labeling. An SRL classifier determines the roles of arguments relative to a predicate in the sentence. The identity of arguments and predicates, of course, is not apparent in the surface form of the sentence. Therefore we suppose that this identification is part of a hidden structure for the sentence. The validity of this hidden structure determines the success of the semantic role labeling. As one of our assumptions about the starting-point of multi-word sentence comprehension, the semantic-role classifier assumes that nouns fill argument slots relative to verbs. Therefore the hidden structure that our system attempts to identify is a simple syntactic structure defined by identifying the nouns and verbs in the sentence, along with their linear order.

As shown in the bottom half of Figure 1, the Latent BabySRL architecture roughly follows the standard pipeline, except that instead of a previously-trained syntactic parser and supervised argument identifier, we rely on an unsupervised clustering of words provided by a Hidden Markov Model (HMM), and a latent argument and predicate identifier that learns in response to feedback from the role classifier. In this system, decisions in the syntactic and semantic layers are linked together, and both are driven by semantic feedback from the world, given appropriate bottom-up information. The latent predicate and argument classifier learns what assists it in predicting semantic roles.

A similar HMM and the experiments in the next section were first presented in (Connor et al., 2010), and a preliminary version of the Latent BabySRL architecture first appeared in (Connor et al., 2011).

3.2.1 Unsupervised Part of Speech Clustering

As a first step in learning we used an unsupervised Hidden Markov Model (HMM) tagger to provide a context-sensitive clustering of words. We fed the learner large

amounts of unlabeled text and allowed it to learn a structure over these data to ground future processing. This stage represents the assumption that the child is naturally exposed to large amounts of language, and will begin to gather distributional statistics over the input, independent of understanding the meaning of any words or sentences. Because we used transcribed speech, this step assumes that the learner can already correctly segment speech into words. The broader sample of text used to support this initial unsupervised HMM clustering came from child directed speech available in the CHILDES repository[3]. We again used only parents' sentences, and we removed sentences with fewer than three words or containing markers of disfluency. In the end we used 320,000 sentences from this set, including over 2 million word tokens and 17,000 unique words. Note that this larger HMM training set included the semantically tagged training data, treated for this purpose as unlabeled text.

The goal of this clustering was to provide a representation that allowed the learner to generalize over word forms. We chose an HMM because an HMM models the input word sequences as resulting from a partially predictable sequence of hidden states. As noted in section 2, distributional statistics over word-strings yield considerable information about grammatical category membership; the HMM states therefore yield a useful unsupervised POS clustering of the input words, based on sequential distributional information, but without names for states. An HMM trained with expectation maximization (EM) is analogous to a simple process of predicting the next word in a stream and correcting connections accordingly for each sentence. We will refer to this HMM system as the HMM 'parser', even though of course parsing involves much more than part-of-speech clustering, largely because in the current version of the Latent BabySRL, the HMM-based clustering fills (part of) the role of the parser in the traditional SRL pipeline shown in Figure 1.

An HMM can also easily incorporate additional knowledge during parameter estimation. The first (and simplest) HMM-based 'parser' we used was an HMM trained using EM with 80 hidden states. The number of hidden states was made relatively large to increase the likelihood of clusters corresponding to a single part of speech, while preserving some degree of generalization. Other researchers (Huang and Yates, 2009) have also found 80 states to be an effective point for creating a representation that is useful for further classification tasks, trading off complexity of training with specificity.

Johnson (2007) observed that EM tends to create word clusters of uniform size, which does not reflect the way words cluster into parts of speech in natural languages. The addition of priors biasing the system toward a skewed allocation of words to classes can help. The second parser we used was an 80-state HMM trained with Variational Bayes EM (VB) incorporating Dirichlet priors (Beal, 2003).[4] These priors assume one simple kind of innate knowledge on the learner's part, represent-

[3] We used parts of the Bloom (Bloom, 1970, 1973), Brent (Brent and Siskind, 2001), Brown (Brown, 1973), Clark (Clark, 1978), Cornell, MacWhinney (MacWhinney, 2000), Post (Demetras et al., 1986) and Providence (Demuth et al., 2006) collections.

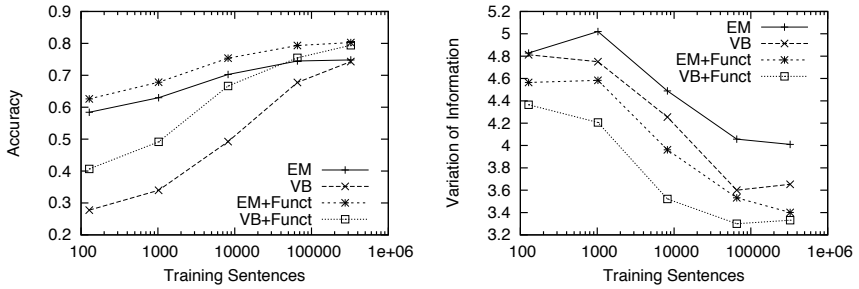
[4] We tuned the priors using the same set of 8 value pairs suggested by Gao and Johnson (2008), using a held out set of POS-tagged CDS to evaluate final performance. Our final values are an emis-

ing the expectation that the language will have a skewed distribution of word classes, with a relatively small number of large classes, and a larger number of small classes.

In the third and fourth parsers we experimented with enriching the HMM with other psycholinguistically plausible knowledge. Words of different grammatical categories differ in their phonological as well as in their distributional properties (e.g., Kelly (1992); Monaghan et al. (2005); Shi et al. (1998)); thus combining phonological and distributional information improves the clustering of words into grammatical categories. The phonological difference between content and function words is particularly striking (Shi et al., 1998). Even newborns can categorically distinguish content versus function words, based on the phonological difference between the two classes (Shi et al., 1999), and toddlers can use both phonology and frequency to identify novel words as likely content versus function words (Hochmann et al., 2010). Human learners may treat content and function words as distinct classes from the start.

To implement this division into function and content words[5], we started with a list of function word POS tags[6] and then found words that appeared predominantly with these POS tags, using tagged WSJ data (Marcus et al., 1993). We allocated a fixed number of states for these function words, and left the rest of the states for the content words. This amounts to initializing the emission matrix for the HMM with a block structure; words from one class cannot be emitted by states allocated to other classes. In previous work (Connor et al., 2010) we selected the exact allocation of states through tuning the heuristic system for argument and predicate identification examined in that work on a held-out set of CDS, settling on 5 states for punctuation, 30 states for function words, and 45 content word states. A similar block structure has been used before in speech recognition work (Rabiner, 1989), and this tactic requires far fewer resources than the full tagging dictionary that is often used to intelligently initialize an unsupervised POS classifier (e.g. Brill (1997); Toutanova and Johnson (2007); Ravi and Knight (2009)).

Because the function versus content word preclustering preceded HMM parameter estimation, it can be combined with either EM or VB learning. Although the initial preclustering independently forces sparsity on the emission matrix and allows more uniform sized clusters within each subset of HMM states, Dirichlet priors may still help, if word clusters within the function or content word subsets vary in size and frequency. Thus the third parser was an 80-state HMM trained with EM estimation, with 30 states pre-allocated to function words; the fourth parser was the same except that it was trained with VB EM.



(a) Many to 1 Accuracy (b) Variation of Information

Fig. 2 Unsupervised Part of Speech results, matching states to gold-standard POS labels. All systems use 80 states, and are evaluated on a POS-labeled subset of CDS text, which comprises a subset of the HMM training data. Many-to-1 matching accuracy greedily matches states to their most frequent part of speech (figure 2(a), higher is better). Variation of Information (figure 2(b)) is an information-theoretic measure summing mutual information between tags and states, proposed by Meilă (2002), and first used for Unsupervised Part of Speech in Goldwater and Griffiths (2007). Smaller numbers are better, indicating less information lost in moving from the HMM states to the gold POS tags. Note that incorporating function word preclustering allowed both EM and VB algorithms to achieve the same performance with an order of magnitude fewer sentences. Figure reproduced from Connor et al. (2010).

3.2.2 HMM Evaluation

In previous work (Connor et al., 2010) we evaluated versions of these parsers (the first stage of our SRL system) on unsupervised POS clustering accuracy. Figure 2 shows the performance of the four parsers described above, using both many-to-one accuracy and variation of information to measure the match between fine-grained POS and the unsupervised parsers' decisions while varying the amount of text they were trained on. Each point on the graph represents the average result over 10 training runs of the HMM with different samples of the unlabeled CDS[7].

Many-to-one accuracy is an evaluation metric that permits multiple HMM states to map onto each POS tag: accuracy is measured by greedily mapping each state to the POS tag it most frequently occurs within the test data; all other occurrences of that state are then considered incorrect. EM can yield a better many-to-one score

sion prior of 0.1 and a transitions prior of 0.0001; as a Dirichlet prior approaches 0 the resulting multinomial becomes peakier with most of the probability mass concentrated in a few points.

[5] We also include a small third class for punctuation, which is discarded.

[6] TO,IN,EX,POS,WDT,PDT,WRB,MD,CC,DT,RP,UH

[7] Note that the data shown in Figure 2 reflect HMM initialization and training that differed slightly from that described in Section 3.2.1 and used in the experiments reported here: In that previous work, the set of function words differed slightly (e.g., in the current version we added 'not' to the function word set, and removed 'like' and 'have'), fewer states were allocated to punctuation (3 rather than 5), and the HMM was trained on a smaller sample of unlabeled text (up to 160,000 sentences rather than 320,000). The revised HMM parser used in the present experiments produced very similar results.

than VB-trained HMM (Johnson, 2007), and our work showed the same result: across variations in amount of training data, EM yielded higher accuracy by this metric than VB, although these distinctions diminished as the amount of training data increased.

Variation of information is a metric of the distance between two clustering solutions (true POS labels and HMM states), which measures the loss and gain of information when moving from one clustering to the other. It is defined as $VI(C_1,C_2) = H(C_1|C_2) + H(C_2|C_1) = H(C_1) + H(C_2) - 2 * I(C_1,C_2)$, where $H(C)$ is the entropy of the clustering assignment C and $I(C_1,C_2)$ is the mutual information between the clustering C_1 and C_2. VI is a valid metric, and thus if two clusterings are identical, their VI will be 0.

These data show that the HMM yielded robust POS clustering, and that the four versions differed from each other in interesting ways. In particular, the content vs. function-word split improved POS clustering performance. Measured both by many-to-1 accuracy and VI, adding the function word split improved performance, for both EM and VB training. Thus a preclustering of content and function words, which we have argued is plausible for learners based on the well-established phonological differences between these classes, improves the automatic identification of POS clusters from text. In future sections we use the VB+Funct HMM, the best-performing system in this evaluation, as the first step in the Latent BabySRL. The HMM states both yield additional representational features that permit generalization across words, and give us a means of incorporating some minimally-supervised syntactic constraints on sentence interpretation.

4 Latent Training

Once a potential predicate and arguments have been identified (via latent training as described in this section), a role classifier must assign a semantic role to each argument relative to the predicate. The role classifier can only rely on features that can be computed with information available from previous stages of input processing, and from prior learning. The latent argument and predicate identifier is trained to best support accurate role classification. We trained this model in an online fashion in which we present each sentence along with some semantic constraints as feedback; both the semantic-role and the latent argument and predicate classifier then update themselves accordingly. In this section we will describe how the model is trained and what representations are used.

We can phrase our problem of Semantic Role Labeling as learning a structured prediction task, which depends on some latent structure (argument and predicate identification). As input we have the sequence of words and HMM states for a given sentence, and the output is a role-labeled predicate-argument structure. The goal in our structured prediction task is to learn a linear function $f_w : X \rightarrow Y$ that maps from the input space X (sentences) to output space Y (role labeled argument structure):

$$f_w(x) = \arg\max_{y \in Y} \max_{h \in H} w \cdot \Phi(x,h,y) \quad (1)$$

Here H is a space of hidden latent structures that describes some connection between X and Y (identification of arguments and predicate), Φ is a feature encoding for the complete role labeled X,H,Y example structure, w is the learned weight vector that scores structures based on their feature encoding, and both $w, \Phi \in \mathbb{R}^n$.

Conventionally the weight vector w would be learned from a set of labeled training examples $(x_i, y_i) \in X \times Y$, attempting to maximize the difference between the score for true structures y_i and all other structures for every training example. As we argued in the introduction to this chapter, it is implausible for the learner to receive veridical sentence meanings for each sentence (the set of role labels linked with arguments) as feedback for learning. The referential contexts that accompany speech are assumed to be ambiguous; this is the problem that syntactic bootstrapping sets out to solve. Therefore, instead of assuming that the learner is provided with a single true interpretation, we rephrase the learning problem such that for each sentence the learner is provided with a set of possible interpretations $Y_i \subseteq Y$ along with constraints on possible hidden structures $H_i \subseteq H$. In the next section we will describe specific implementations of this feedback scheme. However, in this section, for clarity in describing our algorithm and feature-sets, we use as an example the simplest case, in which only the true interpretation is provided.

Because of the max over H in the definition of f_w, the general optimization problem for finding the best w (in terms of minimizing a loss, or maximizing the margin between the true structure and all others given a training set of $\{x_i, y_i\}_{i=1}^M$ labeled examples) is non-convex. Previously this has been solved using some variant of latent structure optimization (Chang et al., 2010; Yu and Joachims, 2009). Here we used an online approach and a modification of Collin's Structured Perceptron (Collins, 2002) with margin (Kazama and Torisawa, 2007). This basic, purely latent algorithm (Algorithm 1) uses an approximation employed in (Felzenszwalb et al., 2008; Cherry and Quirk, 2008) where for each example the best h^* is found (according to the current model and true output structure) and then the classifier is updated using that fixed structure. In this algorithm C is a fixed margin (set at 1.0) that must seperate the true structure from the next highest prediction for the algorithm to not modify the weight vector ($\mathbf{1}[y \neq y_i^*]$ is an indicator function that is 1 for all y that are not the true structure). The constant α_w represents the learning rate.

Algorithm 1 Purely Latent Structure Perceptron

Initialize $w_0, t = 0$
repeat
  for all Sentences (x_i, Y_i) **do**
    $(h_i^*, y_i^*) \leftarrow \arg\max_{h \in H_i, y \in Y_i} w_t \cdot \Phi_w(x_i, h, y)$
    $y' \leftarrow \arg\max_y w_t \cdot \Phi_w(x_i, h_i^*, y) + C * \mathbf{1}[y \neq y_i^*]$
    $w_{t+1} \leftarrow w_t + \alpha_w(\Phi_w(x_i, h_i^*, y_i^*) - \Phi_w(x_i, h_i^*, y'))$
    $t \leftarrow t + 1$
  end for
until Convergence

The intuition behind algorithm 1 is that for every sentence the learner knows the true meaning, or set of meanings that contain the true meaning (Y_i), so it is able to find the arrangement of arguments and predicate (hidden structure h^*) that best supports that meaning according to what it has already learned (current weight vector w_t). Once we identify the latent arguments and predicate, we use this identification to update the weight vector so the true role prediction y_i^* will be more likely in the future (line 5 and 6, structured perceptron update).

As stated in algorithm 1, h^*, the best set of arguments and predicates, is found and then forgotten for each input sentence x. If we are interested in h beyond its application to learning the weights w to predict semantic roles y, such as for generalizing to better find the arguments and predicate in related sentences x, then we need a method for storing this information and passing it on to new examples.

To solve this problem, we trained a latent predicate and argument classifier along with the role classifier, such that during the latent prediction for each sentence we find the structure that maximizes the score of both role classification and structure prediction. This algorithm is summarized in algorithm 2. The end result is two classifiers, f_u to predict hidden structure and f_w to use the hidden structure, that have been trained to work together to minimize semantic-role classification training error.

Algorithm 2 Online Latent Classifier Training

1: Initialize $w_0, u_0, t = 0$
2: **repeat**
3: **for all** Sentences (x_i, Y_i) **do**
4: $(h_i^*, y_i^*) \leftarrow \arg\max_{h \in H_i, y \in Y_i} w_t \cdot \Phi_w(x_i, h, y) + u_t \cdot \Phi_u(x_i, h)$
// Update u to predict h^*
5: $h' \leftarrow \arg\max_h u_t \cdot \Phi_u(x_i, h) + C * \mathbf{1}[h \neq h_i^*]$
6: $u_{t+1} \leftarrow u_t + \alpha_u(\Phi_u(x_i, h_i^*) - \Phi_u(x_i, h'))$
// Update w based on h^* to predict y^*
7: $y' \leftarrow \arg\max_y w_t \cdot \Phi_w(x_i, h_i^*, y) + C * \mathbf{1}[y \neq y_i^*]$
8: $w_{t+1} \leftarrow w_t + \alpha_w(\Phi_w(x_i, h_i^*, y_i^*) - \Phi_w(x_i, h_i^*, y'))$
9: $t \leftarrow t + 1$
10: **end for**
11: **until** Convergence

The intuition behind algorithm 2 is that for each sentence the learner finds the best joint meaning and structure based on the current classifiers and semantic constraints (line 4), then seperately updates the latent structure f_u and output structure f_w classifiers given this selection. In the case where we have perfect high level semantic feedback $Y_i = y_i$, the role classifier will search for the argument structure that is most useful in predicting the correct labels. More generally, partial feedback, which constrains the set of possible interpretations but does not indicate the one true meaning, may be provided and used for both labeling Y_i and hidden structure H_i.

This learning model allows us to experiment with the trade-offs among different possible sources of information for language acquisition. Given perfect or highly informative semantic feedback, our constrained learner can fairly directly infer the true argument(s) for each sentence, and use this as feedback to train the latent argu-

ment and predicate identification (what we might term semantic bootstrapping). On the other hand, if the semantic role feedback is loosened considerably so as *not* to provide information about the true number or identity of arguments in the sentence, the system cannot learn in the same way. In this case, however, the system may still learn if further constraints on the hidden syntactic structure are provided through another route, via a straight-forward implementation of the structure-mapping mechanism for early syntactic bootstrapping.

4.1 Argument, Predicate and Role Classification

For the latent structure training method to work, and for the hidden structure classifier to learn, the semantic role classifier and feature set (f_w and Φ_w respectively) must make use of the hidden structure information h. In our case, the role classifier makes use of (and thus modifies during training) the hidden argument and predicate identification in two ways. The first of these is quite direct: semantic role predictions are made relative to specific arguments and predicates. Semantic-role feedback therefore provides information about the identity of the nouns in the sentence. The second way in which the role classifier makes use of the hidden argument and predicate structure is less direct: The representations used by the SRL classifier determine which aspects of the predictions of the argument and predicate latent classifier are particularly useful in semantic role labeling, and therefore change via the learning permitted by indirect semantic-role feedback.

In the simplest case we use the full set of correct role labels as feedback. We implement this by providing correct labels for each word in the input sentence that was selected by the latent classifier as an argument and is the head noun of an argument-phrase. Thus the optimal prediction by the argument classifier will come to include at least those words. The predicate classifier will therefore learn to identify predicates so as to maximize the accuracy of SRL predictions given these arguments. This represents the case of semantically-driven learning where veridical semantic feedback provides enough information to drive learning of both semantics and syntax. With more ambiguous semantic feedback, the hidden argument and predicate prediction is not directed by straightforward matching of a full set of noun arguments identified via semantic feedback. Nonetheless, the system is still driven to select a hidden structure that best allows the role classifier to predict with what little semantic constraint is provided. Without further constraints on the hidden structure itself, there may not be enough information to drive hidden structure learning.

In turn, the hidden structure prediction of arguments and predicate depends on the words and HMM states below it, both in terms of features for prediction and constraints on possible structures. The hidden argument and predicate structure we are interested in labels each word in the sentence as either an argument (noun), a predicate (verb), or neither. We used the function/content word state split in the HMM to limit prediction of arguments and predicates to only those words identified as content words. In generating the range of possible hidden structures over content

words, the latent structure classifier considers only those with exactly one predicate and one to four arguments.

(a) Sentence Full Feedback	She likes yellow flowers . A0 A1					
(b) Possible Interpretation 1				Possible Interpretation 2		
Sentence Argument Struct.	she N	likes V	yellow flowers N	Sentence Argument Struct.	she N	likes yellow flowers V N
(c) Feature Representation				Feature Representation		
Semantic Feat. $\Phi_w(x,h,y)$	she	argument:she predicate:likes NPat: 1 of 2 VPos:Before Verb w+1:likes		Semantic Feat. $\Phi_w(x,h,y)$	she	argument:she predicate:yellow NPat: 1 of 2 VPos:Before Verb w+1:likes
	flowers	argument:flowers predicate:likes NPat: 2 of 2 VPos: After Verb w-1:yellow w+1:.			flowers	argument:flowers predicate:yellow NPat: 2 of 2 VPos: After Verb w-1:yellow w+1:.
Structure Feat. $\Phi_u(x,h)$	she=N	word:she hmm:35 verb:likes w+1:likes hmm+1:42 NPat: 1 of 2		Structure Feat. $\Phi_u(x,h)$	she=N	word:she hmm:35 verb:yellow w+1:likes hmm+1:42 NPat: 1 of 2
	likes=V	verb:likes hmm:42 w-1:she hmm-1:35 w+1:yellow hmm+1:57 v:likes&2 args suffixes: s,es,kes			yellow=V	verb:yellow hmm:57 w-1:likes hmm-1:42 w+1:flowers hmm+1:37 v:flowers&2 args suffixes: w,ow,low

Table 3 Example Sentence, showing (a) the full (gold standard) semantic feedback that provides true roles for each argument, but no indication of the predicate, as well as (b) two possible hidden structures given this level of feedback. The next rows show (c) the feature representations for individual words. The Semantic Feature set shows the feature representation of each argument as used in SRL classification; the Structure Feature set shows the feature representation of the first argument and the predicate in two of the 28 possible hidden structures. See text section section 4.1.1 for further description of the features.

As an example take the sentence "She likes yellow flowers." There are four content words; with the constraint that exactly one is a predicate and at least one is an argument, there are 28 possible predicate/argument structures, including the correct assignment where 'She' and 'flowers' are arguments of the predicate 'likes.' The full semantic feedback would indicate that 'She' is an agent and 'flowers' is a patient, so the latent score the SRL classifier predicts (line 4 in algorithm 2) will be the sum of the score of assigning agent to 'She' and patient to 'flowers', assuming both those words are selected as arguments in h. If a word does not have a seman-

tic role (such as non-argument-nouns 'likes' or 'yellow' here) then its predictions do not contribute to the score. Through this mechanism the full semantic feedback strongly constrains the latent argument structure to select the true argument nouns. Table 3 shows the two possible interpretations for "She likes yellow flowers." given full semantic feedback that identifies the roles of the correct arguments. Decisions regarding 'likes' and 'yellow' must then depend on the representation used by both the latent-structure predicate identifier and semantic-role classifier.

4.1.1 Features

For the semantic-role classifier we started with the same base BabySRL features developed in Connor et al. (2008), simple structures that can be derived from a linear sequence of candidate nouns and verb. These features include 'noun pattern' features indicating the position of each proposed noun in the ordered set of nouns identified in the sentence (e.g., first of three, second of two, etc; NPat in Table 3), and 'verb position' features indicating the position of each proposed noun relative to the proposed verb (before or after; VPos in Table 3). In the above example, given the correct argument assignment, these features would specify that 'She' is the first of two nouns and 'flowers' is the second of two. No matter whether 'likes' or 'yellow' is selected as a predicate, 'She' is before the verb and 'flowers' is after it. In addition, we used a more complicated feature set that includes NPat and VPos features along with commonly-used features such as the words surrounding each proposed noun argument, and conjunctions of NPat and VPos features with the identified predicate (e.g., the proposed predicate is 'likes' and the target noun is before the verb); such features should make the role classifier more dependent on correct predicate identification.

For the argument and predicate structure classifiers the representation $\Phi_u(x,h)$ only depends on words and the other arguments and predicates in the proposed structure. Each word is represented by its word form, the most likely HMM state given the entire sentence, and the word before and after. We also specified additional features specific to argument or predicate classification: the argument classifier uses noun pattern (NPat in Table 3), and the predicate representation uses the conjunction of the verb and number of arguments (e.g., 'v:likes & 2args' in Table 3), as well as all suffixes of length up to three as a simple verb ending feature[8].

It should be noted that both the purely latent (algorithm 1) and latent classifier we have been discussing (algorithm 2) require finding the max over hidden structures and labelings according to some set of constraints. As implemented with the sentences found in our child directed speech sample, it is possible to search over all possible argument and predicate structures. In our set of training sentences there were at most nine content words in any one sentence, which requires searching over 1458 structures of exactly one predicate and at most four arguments. On average

[8] This roughly represents phonological/distribution information that might be useful for clustering verbs together (e.g., Monaghan et al. (2005)), but that is not exploited by our HMM because the HMM takes transcribed words as input.

there were only 3.5 content words a sentence. Once we move to more complicated language an alternative approximate search strategy will need to be employed.

5 Experimental Evaluation

To evaluate the Latent BabySRL, we examined both how well the final role classifier performed, and how accurately the latent predicate and argument classifiers identified the correct structures when trained with only indirect semantic feedback. Because in our training sentences there was only one true predicate per sentence, we report the predicate accuracy as the percentage of sentences with the correct predicate identified. For the identification of noun arguments, because there were multiple possible predictions per sentence, we report F1: the harmonic mean of precision and recall in identifying true arguments. Likewise, in evaluating semantic-role classification, because there were many possible role labels and arguments to be labeled, we report the overall semantic role F1 over all arguments and label predictions[9].

Our first experiment tested online latent training with *full semantic feedback*. To provide an upper bound comparison we trained with perfect argument knowledge, so in this case both classifiers were fully and separately supervised (Gold Arguments in Table 4). This upper bound reflects the levels of argument-identification and SRL performance that are possible given our simple feature set and child-directed sentence corpus. As a lower bound comparison for predicate-argument classification we also include the expected result of selecting a random predicate/argument structure for each sentence (Random Arguments in Table 4).

Training	Predicate %	Argument F1	Role F1
Gold Arguments	0.9740	0.9238	0.6920
Purely Latent	0.5844	0.6992	0.5588
Latent Classifier	0.9263	0.8619	0.6623
Random Arguments	0.3126	0.4580	-

Table 4 Results on held-out test set of SRL with arguments/predicate as latent structure, provided with full semantic feedback. With Gold Arguments, both the structure classifier and the role classifier are trained with full knowledge of the correct arguments for each sentence. Purely Latent does not use the latent argument and predicate classifier; it selects a structure for each sentence that maximizes role classification of true labels during training (algorithm 1). Latent Classifier training trains an argument/predicate identifier using the structure that the role classifier considers most likely to give the correct labeling (where we know correct labels for each noun argument), algorithm 2.

Table 4 shows the performance of the two algorithms from section 4 compared to the just-mentioned upper and lower bounds. All classifiers used the full feature sets from section 4.1. Recall that the purely latent method (algorithm 1) did not use an

[9] Because we focus on noun arguments, we miss those predicate arguments that do not include any nouns; the maximum SRL role F1 with only noun arguments correct is 0.8255.

intermediate latent structure classifier, so it selected arguments and predicates only to maximize the role classifier prediction for the current sentence. In contrast, incorporating a latent classifier into the training (algorithm 2) yielded a large boost in both argument and predicate identification performance and final role performance. Thus, given full semantic feedback, the argument and predicate classifier effectively generalized the training signal provided by the latent semantic feedback to achieve nearly the performance of being trained on the true arguments explicitly (Gold Arguments). Of special note is the predicate identification performance; while full semantic feedback implicitly indicates true arguments, it says nothing about the true predicates. The predicate classifier was able to extract this information solely based on identifying latent structures that helped the role classifier make the correct role predictions.

As mentioned in section 4.1, our algorithm depends on two kinds of representations: those that feed semantic role classification, and those that feed the hidden argument and predicate classifier. To investigate the interaction between the two classifiers' (hidden structure and SRL) representation choices, we tested the latent classifier with the full argument and predicate feature sets when the role classifier incorporated four different feature sets of increasing complexity: only the words identified as candidate nouns and verb (Words in Table 5), words plus noun pattern features (+NPat), the previous plus verb position features (+VPos), and a full model containing all these features as well as surrounding words and predicate conjunctions. With the addition to the SRL classifier of features that depend on more accurate latent structure identification, we should see improvements in both final role accuracy and argument and predicate identification. This experiment again used full role feedback.

Role Features	Predicate %	Argument F1	Role F1
Words	0.64 (0.02)	0.81 (0.00)	0.63 (0.01)
+NPat	0.73 (0.05)	0.81 (0.00)	0.62 (0.01)
+VPos	0.93 (0.04)	0.83 (0.03)	0.65 (0.01)
+Surrounding words and predicate conjunctions	0.93 (0.03)	0.86 (0.04)	0.66 (0.01)

Table 5 With full role feedback and latent classifier training, the role classifier features interact with the latent predicate-argument structure classifier. Better role classification through improved feature representation feeds back to allow for improved argument and predicate identification. The last two feature sets make strong use of the identity of the predicate, which encourages the predicate classifier to accurately identify the predicate. Each result represents the average over ten runs with random training order; numbers in parenthesis are standard deviations.

Table 5 shows increasing performance with the increasing feature complexity of the semantic role classifier. Most notable is the large difference in predicate identification performance between those feature sets that heavily depend on accurate predicate information (+VPos and the full feature set in Table 5) and those that only use the word form of the identified predicate as a feature. In contrast, argument identification performance varied much less across feature sets in this experiment,

because full semantic feedback always implicitly drives accurate argument identification. The increase in role classification performance across feature sets can be attributed both to a useful increase in representations used for SRL classification, and to the increased argument and predicate structure accuracy during both SRL training and testing. The relatively high level of SRL performance given the lexical features alone in Table 5 reflects the repetitive character of the corpus from which our training and test sentences were drawn: Given full semantic feedback, considerable success in role assignment can be achieved based on the argument-role biases of the target nouns (e.g., 'she', 'flowers') and the familiar verbs in our corpus of child-directed speech.

The results in this section show that the latent argument and predicate classifier, equipped with simple representations of the proposed sentence structure, can recruit indirect semantic-role feedback to learn to improve its representation of sentence structure, at least when given fully accurate semantic-role feedback. This result makes sense: the identity and position of the verb are useful in identifying the semantic roles of the verb's arguments; therefore the latent predicate-argument classifier could use the indirect semantic feedback to determine which word in the sentence was the verb. The full semantic-role feedback provided true information about the number and identity of arguments in each sentence; in the next section we take the crucial next step, reducing the integrity of the semantic role feedback to better reflect real-world ambiguity.

5.1 Ambiguous Semantic Feedback

The full semantic feedback used in the previous experiments, while less informative than absolute gold knowledge of true arguments and predicates, is still an unreasonable amount of feedback to grant a child first trying to understand sentences. The semantic feedback in our model represents the child's inference of sentence meaning from observation of the referential context. Because an ordinary scene makes available a number of possible objects, relations and semantic roles that might be mentioned, the child must learn to interpret sentences without prior knowledge of the true argument labels for the sentence or of how many arguments are present.

We implement this level of feedback by modifying the constraining sets H_i and Y_i used in line 4 of algorithm 2. By loosening these sets we still provide feedback to restrict the search space (thus modeling language learning as a partially supervised task, informed by inferences about meaning from scene observation), but not a veridical role-labeling for each sentence.

We tested two levels of reduced role feedback. The first, which we call Set of Labels, provides as feedback the true role labels that are present in the sentence, but does not indicate which words correspond to each role. In this case Y_i is just the set of all labelings that use exactly the true labels present, and H_i is constrained to be only those syntactic predicate-argument structures with the correct number of arguments. This feedback scheme represents a setting where the child knows

the semantic relation involved, but either does not know the nouns in the sentence, or alternatively does not know whether the speaker meant 'chase' or 'flee' (and therefore cannot fix role order). To illustrate, given the sentence "Sarah chased Bill", Set of Labels feedback would indicate only that the sentence's meaning contains an agent and a patient, but not which word in the sentence plays which role.

Even this Set of Labels feedback scheme specifies the number of true arguments in the sentence. We can go a step further, supplying for each sentence a superset of the true labels from which the learner must select a labeling. In the Superset feedback case, Y_i includes the true labels, plus random additional labels such that for every sentence there are 4 labels to choose from, no matter the number of true arguments. Given Superset feedback, the learner is no longer constrained by the true number of arguments provided via semantic feedback, so must search over all argument structures and role labelings that come from some subset of the feedback set Y_i. This represents a setting in which the learner must select a possible interpretation of the sentence from a superset of possible meanings provided by the world around them. In the "Sarah chased Bill" example, the feedback would be a set of possible labels including the true agent and patient roles, but also two other roles such as recipient or location, and thus no scene-derived indication of how many of these roles are part of the sentence's meaning. This may seem an extreme reduction of the validity of semantic-role feedback. However, consider the following example: a careful analysis of video transcripts of parents talking to toddlers found that the parents were about equally likely to use intransitive motion verbs (e.g., 'go in') as transitive ones (e.g., 'put in') when describing events in which an agent acted on an object (Rispoli, 1989). Evidently the presence of an agent in an event does not demand that a speaker choose a verb that encodes the agent's role. Similarly, in our earlier 'yellow flower' example, under many circumstances the speaker presumably could have said 'yellow flowers are nice' rather than 'she likes yellow flowers.' These considerations, and the 'human simulation' experiments described in section 1.1 (Gillette et al., 1999), all suggest that the number and identity of arguments in the speaker's intended meaning is not readily inferrable from world events without some guidance from the sentence.

Feedback	Pred %	Arg F1	A0	A1	Role F1
Full Labels	0.94 (0.02)	0.89 (0.02)	0.85 (0.02)	0.75 (0.02)	0.64 (0.02)
Set of Labels	0.40 (0.23)	0.62 (0.14)	0.47 (0.28)	0.38 (0.17)	0.34 (0.14)
Superset	0.35 (0.20)	0.57 (0.11)	0.46 (0.27)	0.33 (0.13)	0.29 (0.11)
Random	0.31	0.46			

Table 6 Results when the amount of semantic feedback is decreased. Each value represents the mean over twenty training runs with shuffled sentence order; the numbers in parenthesis are the standard deviations. Full label feedback provides true role feedback for each noun. Set of Labels feedback provides an unordered set of true labels as feedback, so the learner must pick a structure and label assignment from this set. Superset goes one step further and provides a superset of labels that includes the true labels, so the learner does not know how many or which roles are mentioned in the sentence. With these ambiguous feedback schemes the classifiers are barely able to begin interpreting correctly, and with superset feedback the argument and predicate accuracy is only slightly better than random.

As seen in Table 6, Set and Superset feedback seriously degrade performance compared to full role feedback. With superset feedback the learner cannot get a good foothold to begin correctly identifying structure and interpreting sentences, so its argument and predicate identification accuracy is little better than random. This suggests that information about the number and identity of arguments might be a necessary constraint in learning to understand sentences. In principle this information could be derived either from observation of scenes (assuming the child has access to some non-linguistic source of evidence about whether the speaker meant 'chase' or 'flee', 'put in' or 'go in') or from observation of sentences; the latter source of information is the essence of syntactic bootstrapping, as we discuss next.

6 Recovering Argument Knowledge

Considerable psycholinguistic evidence, reviewed briefly in Section 1.1, suggests that children learn some nouns before they start to interpret multi-word sentences, and thus some noun knowledge is available to scaffold the beginnings of sentence interpretation (e.g., Gillette et al. (1999)). This is syntactic bootstrapping, using structural features of the sentence to guide interpretation under ambiguity. If we can combine this extra source of knowledge with the Superset feedback described above, then perhaps the result will be enough information for the system to learn to identify nouns and verbs in sentences, and to classify the roles those nouns play.

Taking inspiration from the 'structure-mapping' account of syntactic bootstrapping, we model this starting point by attempting to identify nouns in each input sentence in a bottom-up, minimally-supervised manner. Once we know the number and identity of nouns in the sentence, this additional constraint on the hidden structure may allow the learner to overcome the semantic ambiguity introduced by Superset feedback. In the next section we will describe how we identify potential arguments using the distributional clustering provided by the HMM and a small seed set of concrete nouns. A similar form of this bottom-up argument-identification procedure was described in (Connor et al., 2010).

The bottom-up, minimally-supervised argument identifier we describe here addresses two problems facing the human learner. The first involves clustering words by part-of-speech. As described in section 3.2.1, we use a fairly standard Hidden Markov Model (HMM), supplemented by an a priori split between content and function words, to generate clusters of words that occur in similar distributional contexts. The second problem is more contentious: Having identified clusters of distributionally-similar words, how do children figure out what role these clusters of words play in a sentence interpretation system? Some clusters contain nouns, which are candidate arguments; others contain verbs, which take arguments. How is the child to know which are which?

The latent training procedure described in this chapter, when given full semantic feedback, accomplishes argument and predicate identification roughly by semantic bootstrapping: To return to our 'She likes yellow flowers' example, if the learner

knows based on semantic feedback that 'she' is an agent, then the latent classifier learns to treat 'she' as a noun argument; the provision of abstract HMM-based features and noun-pattern features to the argument identification classifier permits it to generalize this learning to other words in similar sentence positions. But this use of semantic-role feedback to identify nouns as such seems counter-intuitive. In this section we spell out a simpler way to use a small step of semantic bootstrapping to automatically label some of the distributionally-derived clusters produced by the HMM tagger as nouns, thereby improving argument-identification from the bottom up, without requiring accurate semantic-role feedback.

6.1 Bottom-Up Argument Identification

The unsupervised HMM parser provides a state label for each word in each sentence; the goal of the argument identification stage is to use these states to label words as potential arguments, predicates or neither. As described in Section 1.1, the structure-mapping account of early syntactic bootstrapping holds that sentence comprehension is grounded in the learning of an initial set of nouns. Children are assumed to identify the referents of some concrete nouns via cross-situational learning (Gillette et al., 1999; Smith and Yu, 2008). Children then assume, given the referential meanings of these nouns, that they are candidate arguments. Again, this involves a small step of semantic bootstrapping, using the referential semantics of already-learned words to identify them as nouns. We used a small set of known nouns to transform unlabeled word clusters into candidate arguments for the SRL: HMM states that occur frequently with known names for animate or inanimate objects are assumed to be argument states.

Algorithm 3 Argument State Identification

INPUT: Parsed Text T = list of (word, state) pairs
Set of concrete nouns N
OUTPUT: Set of argument states A
$A \leftarrow \varnothing$
// Count Appearance of each state with a known noun
$freq_N(s) \leftarrow |\{(w,s) \in T | w \in N\}|$
for all Content States s **do**
  if $freq_N(s) \geq 4$ **then**
    Add s to A
  end if
end for
return A

Given text parsed by the HMM parser and a seed list of known nouns, the argument identifier proceeds as illustrated in algorithm 3. Algorithm 3 identifies noun states simply by counting the number of times each state is seen with a known noun ($freq_N(s)$ in algorithm 3) in some HMM tagged text (Adam training data). Any

state that appears at least 4 times with words from the seed noun list is identified as a noun state. Whenever these states are encountered in the future, the word associated with them, even if unknown, will be interpreted as a potential argument. This use of a seed list with distributional clustering is similar to Prototype Driven Learning (Haghighi and Klein, 2006), except in the present case we provide information on only one class. A similar approach was proposed by Mintz (2003), using semantic knowledge of a small set of seed nouns to tag pre-existing distributionally-based clusters as noun clusters.

Because we train our HMM with a preclustering of states into function and content words, we use this information in the minimally supervised argument identification. Only content word states are considered to be potential argument states, thus eliminating any extraneous function words from consideration. This of course improves identification performance, because it only eliminates potential errors.

To generate a plausible 'seed' set of concrete nouns, we used lexical development norms (Dale and Fenson, 1996), selecting all words for things or people that were commonly produced by 20-month-olds (over 50% reported), and that appeared at least 5 times in our training data. Because this is a list of words that children produce, it represents a lower bound on the set of words that children at this age should comprehend. This yielded 71 words, including words for common animals ('pig', 'kitty', 'puppy'), objects ('truck', 'banana', 'telephone'), people ('mommy', 'daddy'), and some pronouns ('me' and 'mine'). To this set we added the pronouns 'you' and 'I', as well as given names 'adam', 'eve' and 'sarah'. The inclusion of pronouns in our list of known nouns represents the assumption that toddlers have already identified pronouns as referential terms. Even 19-month-olds assign appropriately different interpretations to novel verbs presented in simple transitive versus intransitive sentences with pronoun arguments ("He's kradding him!" vs. "He's kradding!"; (Yuan et al., ress)).

The resulting set of 76 seed nouns represents a high-precision set of argument nouns: They are not highly frequent in the data (except for the pronouns), but they nearly always appear as nouns and as arguments in the data (over 99% of the occurrences of words in this list in our training data are nouns or pronouns, over 97% are part of arguments). Given this high precision, we set a very permissive condition that identifies argument states as those HMM states that appear 4 or more times with known seed nouns. In our experiments we set the threshold of known nouns appearing with an HMM state to 4 through tuning argument identification on a held-out set of argument-identified sentences.

6.1.1 Argument Identification Evaluation

Figure 3 shows the argument identification accuracy of the minimally supervised argument identification system, with increasing numbers of seed nouns sampled from the set of 76. First, using the HMM model with the function-content-word split and VB training, we generated 10 models over the large untagged HMM training corpus with different random initializations, and selected the one with the lowest perplex-

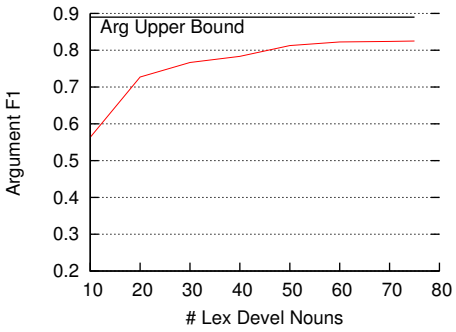


Fig. 3 Effect of number of concrete nouns for seeding argument identification. To generate these results, one HMM trained with VB+Funct was selected out of ten models with different random initializations (the best in terms of lowest perplexity on large corpus of untagged training data). Adam training data was then tagged with states from this HMM; for each lexical development seed set size, 100 runs with random selection of seed nouns were averaged together to produce the data shown here. Argument identification accuracy is computed for the Adam data set using true argument boundaries from hand labeled data. With 76 seed nouns, the argument identifier achieves over 0.80 F1.

ity (highest log likelihood) for use in the argument identification experiments. We then tagged the Adam SRL training data with states from this selected HMM, using the state for each word that had highest marginal probability given the rest of the sentence (using forward-backward algorithm). In this figure, for each set size of seed nouns, we report the mean over 100 runs of the argument identification with a random selection of seed nouns in each run.

We evaluated this performance compared to hand labeled data with true argument and predicate boundaries. We present the primary argument (A0-4) identification accuracy using the F1 measure, with precision calculated as the proportion of identified arguments that appear as part of a true argument, and recall as the proportion of true arguments that cover some state identified as an argument. This is a rather lenient measure of accuracy since we are comparing identified individual words to full phrase boundaries.

As Figure 3 shows, this minimally-supervised argument identification system can successfully identify arguments starting with a handful of concrete nouns. Even with just 10 nouns, argument identification is almost 0.6 F1; with 76 nouns (still a modest number relative to toddlers' estimated comprehension vocabularies), argument identification improves to over 0.8 F1. Even so, we have not yet tapped the full information available in the finer grained HMM clusters. Looking at the upper bound, which is computed as the optimal selection of HMM states given knowledge of true argument boundaries, there is still some room for improvement.

6.2 Integrating into Online Latent Classifier

Next, we use this bottom-up argument identification system to constrain the argument search in our latent classifier training. During training, we restrict the set of possible argument structures (H_i in algorithm 2) such that only those structures that agree with the HMM argument identification are considered, and the best labeling from the Superset of labels is selected for this structure. If we use the arguments identified via HMM argument identification to essentially fix the argument structure during training, the problem remaining for the learner is to select the predicate from among the non-argument content words in the sentence, while also identifying the labeling that is most consistent with the identified arguments and expectations from previous training.

Feedback	Pred %	Arg F1	A0	A1	Role F1
Full Labels	0.94(0.02)	0.89(0.02)	0.85(0.02)	0.75(0.02)	0.64(0.02)
Set of Labels	0.40(0.23)	0.62(0.14)	0.47(0.28)	0.38(0.17)	0.34(0.14)
Superset	0.35(0.20)	0.57(0.11)	0.46(0.27)	0.33(0.13)	0.29(0.11)
Superset + HMM Args	0.87(0.10)	0.88(0.01)	0.68(0.25)	0.54(0.16)	0.48(0.15)
Superset + True Args	0.86(0.09)	0.92(0.01)	0.69(0.21)	0.61(0.13)	0.52(0.13)
Superset + True Args&Pred	0.97(0.00)	0.93(0.00)	0.68(0.19)	0.61(0.12)	0.52(0.11)
Random	0.31	0.46			

Table 7 Results when the amount of semantic feedback is decreased, but bottom-up syntactic information is used to help constrain the possible hidden structures and recover from ambiguous semantic feedback. The top three rows of data are reproduced from Table 6. We introduce extra information by constraining the possible argument structures for each training example using syntactic knowledge, either bottom-up from an HMM-based minimally supervised argument identifier, or via knowledge of true arguments. Once extra information about argument identity is introduced, whether true arguments or the HMM-identified arguments, the learner is able to make use of the Superset feedback, and begin to identify the agent and patient roles (A0 and A1), and the predicate.

Table 7 shows that once we add the HMM bottom-up argument identification to the Superset feedback scheme, the argument and predicate performance increases greatly (due to accuracy of the HMM argument identification). Note in Table 7 that bottom-up HMM argument identification is strong (0.88 F1 compared to 0.93 when trained with true arguments), and that this effective argument-identification in turn permits strong performance on verb identification. Thus our procedure for tagging some HMM classes as argument (noun) classes based on a seed set of concrete nouns, combined with ambiguous Superset semantic feedback that does not indicate the number or identity of semantic arguments, yields enough information to begin learning to identify predicates (verbs) in input sentences.

Next, looking at the final role classification performance of the Superset+argument constraint training schemes in Table 7, we see that Role F1 increases over both straight Superset and unordered Set of Labels feedback schemes. This increase is most dramatic for the more common A0 and A1 roles.

This represents one possible implementation of the structure-mapping procedure for early syntactic bootstrapping. If we assume the learner can learn some nouns

with no guidance from syntactic knowledge (represented by our seed nouns), that noun knowledge can be combined with distributional learning (represented by our HMM parser) to tag some word-classes as noun classes. Representing each sentence as containing some number of these nouns (HMM argument identification) then permits the Latent BabySRL to begin learning to assign semantic roles to those nouns in sentences given highly ambiguous feedback, and also to use that ambiguous semantic feedback, combined with the constraints provided by the set of identified nouns in the sentence, to improve the latent syntactic representation, beginning to identify verbs in sentences.

This latent training method with ambiguous feedback works because it is seeking consistency in the features of the structures it sees. At the start of training, or when encountering a novel sentence with features not seen before, the latent inference will essentially choose a structure and labeling at random (since all structures will have the same score of 0, and ties are broken randomly). From this random labeling the classifier will increase connection strengths between lexical and structural features in the input sentence, and the (at first randomly) selected semantic role labels. Assuming that some number of random or quasi-random predictions are initially made, the learner can only improve if some feature weights increase above the others and begin to dominate predictions, both in the latent structure classifier and in the linked SRL classifier. This dominance can emerge only if there are structural features of sentences that frequently co-occur with frequent semantic roles.

Thus, the assignment of A0 and A1 roles can be learned by this latent SRL learner despite superset feedback, both because of the frequency of these two roles in the training data and their consistent co-occurrence with simple sentence-structure features that make use of the bottom-up information provided by the HMM argument identification. If “She likes yellow flowers.” is encountered early during latent training, the feedback may be the superset {A0, A1, A4, AM-LOC}, where the true labels A0 and A1 are present along with two other random labels. With accurate identification of ‘she’ and ‘flowers’ as arguments via the HMM bottom-up argument identification system, the learner will choose among only those role labelings that use two of the four roles. Given a large number of different sentences such as “She kicks the ball” (true labels are A0, A1), “She writes in her book” (A0, A2), and “She sleeps” (A0), the most consistent labeling amongst the true and random labelings provided by Superset feedback will be that both ‘she’ and the first of two nouns are more likely to be labeled as A0. This consistent labeling is then propogated through the learner’s weights, and used for future predictions and learning. Thus, even superset feedback can be informative given bottom-up information about the *nouns* in the sentence, because frequent nouns and argument patterns (e.g., first of two nouns) consistently co-occur with frequent roles (e.g., A0). Without the identified arguments, the chance of randomly assigning the correct arguments and roles decreases dramatically; as a result, the likelihood of encountering the correct interpretation often enough for it to dominate disappears.

7 Conclusion

We began with two problems for accounts of language acquisition: The sequences of words that make up the input sentences constitute highly ambiguous evidence for syntactic structure, and the situations in the world that accompany the input sentences constitute highly ambiguous evidence for sentence meaning. These two problems have led to 'bootstrapping' approaches to language acquisition, in which some set of built-in representational or architectural constraints on the language-learning system permit the learner to infer one type of structure from knowledge of another. Via semantic bootstrapping (Pinker, 1984, 1989), the learner uses independent knowledge of word and sentence meaning to identify the covert syntactic structures of sentences. Via syntactic bootstrapping (Fisher et al., 2010; Gillette et al., 1999; Landau and Gleitman, 1985; Naigles, 1990), the learner uses independent partial knowledge of syntactic structures to determine sentence meaning. These views are sometimes described as competing accounts, but in fact they share many assumptions, crucially including the assumption that the learner begins with some constraints on the possible links between syntax and semantics. In the present work we tried to incorporate key intuitions of both semantic and syntactic bootstrapping accounts to jointly address the two ambiguity problems with which we began.

To do so, we created a system within which we could manipulate the provision of partially-reliable syntactic and semantic information sources during language acquisition. We trained a semantic role classifier jointly with a simplified latent syntactic structure classifier, with learning based on (varyingly ambiguous) semantic feedback and simple linguistic constraints. This Latent BabySRL, sketched in Figure 1, began by using an HMM to cluster unlabeled word-forms by part of speech. This clustering was based on distributional information recoverable from input word sequences, and was constrained by an initial division into content and function words, and by prior biases regarding the sparsity of word classes. This step represents the assumption that infant learners, even before they understand the meanings of words or sentences, gather statistics about how words are distributed in the linguistic input (e.g., Gomez and Gerken (1999); Marcus. et al. (1999); Romberg and Saffran (2010)), and discriminate content from function words based on their distinct phonological properties (Shi et al., 1998, 1999). It is well established in previous work that considerable information about grammatical category similarity can be obtained by the kind of sequential distributional analysis that an HMM undertakes. We assume that other learning architectures that are sensitive to the sequential statistics of the input would produce similar results; this would include a simple recurrent network that learns a category structure in its hidden units to predict the next word in input sentences (e.g., Chang et al. (2006); Elman (1990)).

With this previous distributional learning in hand, the Latent BabySRL attempted to jointly learn a latent structure for identifying arguments (nouns) and a predicate (verb) in input sentences, and to predict the roles of the identified arguments relative to the identified predicate. The only information sources for this joint learning task were the semantic-role feedback (ranging from full 'gold standard' feedback to highly ambiguous superset feedback) provided to the semantic-role classifier, the

representational constraints on the two classifiers (their feature sets), and the way in which the predictions of the latent structure classifier were used to generate input features for the semantic role classifier. These constraints represent simple but substantive constraints on the links between syntax and semantics. First, the semantic role classifier predicts a semantic role for all and only the nouns it finds in the input sentence. This represents a simple built-in link between syntax and semantics, and a key assumption of the structure-mapping view: the learner assumes each noun is an argument of some predicate term. Second, the latent structure classifier and the semantic-role classifier are equipped with both lexical features (the words in the sentence) and more abstract structural features that permit them to generalize beyond particular words. These abstract features include the predicted HMM clusters of the nouns and verb identified in the sentence, and also simple sentence-structure relational features that can be derived from the identified sequence of nouns and verb, features such as "1st of 2 nouns" and "preverbal noun". Crucially, the specific content and the connection weights of these simple abstract structural features are not provided to the model as hand-coded features of input sentences; such a choice would model a learner that (somehow) begins with accurate identification of nouns and verbs. Instead, the specific syntactic and semantic knowledge that develops in the system arises from the kinds of features the classifiers can represent, and the way in which the model is set up to use them to identify latent structures and in turn to predict semantic roles. Thus we model a learner that begins with substantive constraints on links between syntax and semantics, but without being informed of which words are nouns and which are verbs.

When trained with very informative semantic-role feedback, the Latent BabySRL implements a simple form of semantic bootstrapping. The provision of complete semantic role feedback represents the assumption that the child knows the meaning of the content words in the sentence, and can generate an interpretation of the input sentence based on observing the accompanying scene. Given such veridical semantic feedback, the Latent BabySRL can straightforwardly identify the noun arguments in the sentence (they are the ones that play semantic roles such as agent or patient), but can also learn to identify the verb, by learning that the identity and position of the verb are useful predictor of semantic roles in English sentences (e.g., preverbal nouns tend to be agents).

When trained with highly ambiguous semantic feedback, the Latent BabySRL still learned to identify arguments and predicates, and to use that inferred syntactic structure to assign semantic roles, but only if the system was 'primed' with knowledge of a small set of concrete nouns. The superset feedback described in Section 5.1 made possible many interpretations of each input sentence (including the true one); this feedback scheme provided no information about the number of arguments in each sentence, or which word in the sentence should be aligned with each semantic role. We implemented a procedure whereby a set of concrete seed nouns was used to automatically tag some HMM clusters as noun clusters. This bottom-up argument identification system then constrained the argument search in the latent structure classifier training, as described in Section 6.2. Representing each input sentence as containing some number of nouns guided the learner's assignment of meaning to

input sentences; this in turn permitted the Latent BabySRL to improve its representation of input sentences (including learning to identify the verb), and therefore to further improve its semantic-role classification.

This process represents one straightforward implementation of the structure-mapping account of the origin of syntactic bootstrapping. A skeletal sentence structure, grounded in a set of concrete nouns, provides a preliminary estimate of the number and identity of the noun arguments in the sentence, which in turn permits further semantic and syntactic learning. The Latent BabySRL's dramatic failure to learn when provided with superset feedback without this bottom-up information about the number of noun arguments in the sentence suggests that argument-number information, which in principle could be derived from lucky observations of informative scenes (as in the full-feedback version), or from partial knowledge of syntax grounded in a set of nouns, was crucial to permitting the system to learn.

One might ask which of these two settings of our model is closer to the typical state of the human learner. Should we assume the semantic bootstrapping setting is typical – that the child often knows the meanings of the content words in sentences, and can divine the sentence's meaning from observation of scenes? Or should we assume that the syntactic bootstrapping setting is typical, particularly at early points in acquisition – that the child needs guidance from the sentence itself to determine the abstract relational meanings of verbs, and of sentences? Some would argue that even toddlers can often determine the speaker's intended message in contexts of face-to-face interaction, reading the speaker's intention in a shared interactional goal space (e.g., Pinker (1989); Tomasello (2003)). Others, including the present authors, would argue that the abstract relational meanings of verbs and sentences cannot routinely be determined from event observation without linguistic guidance (e.g., Fisher (1996); Gillette et al. (1999); Rispoli (1989)). The present computational experiments contribute to this conversation by making explicit one way in which partial representations of the structure of sentences, derived with the aid of no semantic information beyond the meanings of a few concrete nouns, to guide early verb learning and sentence interpretation.

Acknowledgements We wish to thank Yael Gertner for insightful discussion that led up to this work as well as the various annotators who helped create the semantically tagged data. This research is supported by NSF grant BCS-0620257 and NIH grant R01-HD054448.

References

Alishahi, A. and Stevenson, S. (2010). A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, 25(1):50–93.

Alishahi, A. and Stevenson, S. (2012). Gradual acquisition of verb selectional prefences in a bayesian model. In Villavicencio, A., Alishahi, A., Poibeau, T.,

and Korhonen, A., editors, *Cognitive Aspects of Computational Language Acquisition*. Springer.

Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In Hayes, J., editor, *Cognition and the development of language*, pages 279–362. John Wiley & Sons, New York, NY.

Bloom, B. H. (1970). Space/Time trade-offs in Hash Coding with allowable errors. *Communications of the ACM*, 13(7):422–426.

Bloom, L. (1973). *One word at a time: The use of single-word utterances before syntax*. Mouton, The Hague.

Bod, R. (2009). From exemplar to grammar: a probabilistic analogy-based model of language learning. *Cognitive Science*, 33(5):752–793.

Brent, M. R. and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81:31–44.

Brill, E. (1997). Unsupervised learning of disambiguation rules for part of speech tagging. In *Natural Language Processing Using Very Large Corpora*. Kluwer Academic Press.

Brown, P., Pietra, V. D., deSouza, P., Lai, J., and Mercer, R. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479.

Brown, R. (1973). *A First Language*. Harvard University Press, Cambridge, MA.

Carreras, X. and Màrquez, L. (2004). Introduction to the CoNLL-2004 shared tasks: Semantic role labeling. In *Proceedings of CoNLL-2004*, pages 89–97. Boston, MA, USA.

Carreras, X. and Màrquez, L. (2005). Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*.

Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychological review*, 113(2):234–72.

Chang, M., Goldwasser, D., Roth, D., and Srikumar, V. (2010). Discriminative learning over constrained latent representations. In *Proc. of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*.

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.

Cherry, C. and Quirk, C. (2008). Discriminative, syntactic language modeling through latent svms. In *Proc. of the Eighth Conference of AMTA*, Honolulu, Hawaii.

Clark, E. V. (1978). Awwareness of language: Some evidence from what children say and do. In Sinclair, R. J. A. and Levelt, W., editors, *The child's conception of language*. Springer Verlag, Berlin.

Clark, E. V. (1990). Speaker perspective in language acquisition. *Linguistics*, 28:1201–1220.

Collins, M. (2002). Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *EMNLP*.

Connor, M., Fisher, C., and Roth, D. (2011). Online latent structure training for language acquisition. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2008). Baby srl: Modeling early language acquisition. In *Proc. of the Annual Conference on Computational Natural Language Learning (CoNLL)*.

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2009). Minimally supervised model of early language acquisition. In *Proc. of the Annual Conference on Computational Natural Language Learning (CoNLL)*.

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2010). Starting from scratch in semantic role labeling. In *Proc. of the Annual Meeting of the Association of Computational Linguistics (ACL)*, Uppsala, Sweden.

Dale, P. S. and Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28:125–127.

Demetras, M., Post, K., and Snow, C. (1986). Feedback to first-language learners. *Journal of Child Language*, 13:275–292.

Demuth, K., Culbertson, J., and Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of english. *Language & Speech*, 49:137–174.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67:547–619.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14:179 –211.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225.

Felzenszwalb, P., McAllester, D., and Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska*.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretation of sentences. *Cognitive Psychology*, 31:41–81.

Fisher, C., Gertner, Y., Scott, R., and Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1:143–149.

Fisher, C., H. Gleitman, H., and Gleitman, L. (1989). On the semantic content of subcategorization frames. *Cognitive Psychology*, 23:331–392.

Fisher, C. and Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, 67:3192–3218.

Gao, J. and Johnson, M. (2008). A comparison of bayesian estimators for unsupervised Hidden Markov Model POS taggers. In *Proceedings of EMNLP-2008*, pages 344–352.

Gentner, D. (2006). Why verbs are hard to learn. In Hirsh-Pasek, K. and Golinkoff, R., editors, *Action meets word: How children learn verbs*, pages 544–564. Oxford University Press.

Gertner, Y., Fisher, C., and Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17.

Gildea, D. and Palmer, M. (2002). The necessity of parsing for predicate argument recognition. In *ACL*, pages 239–246.

Gillette, J., Gleitman, H., Gleitman, L. R., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73:135–176.

Goldwater, S. and Griffiths, T. (2007). A fully bayesian approach to unsupervised part-of-speech tagging. In *ACL*, pages 744–751, Prague, Czech Republic.

Gomez, R. and Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70:109–135.

Haghighi, A. and Klein, D. (2006). Prototype-driven learning for sequence models. In *Proc. of HTL-NAACL*.

Hajič, J., Ciaramita, M., Johansson, R., Kawahara, D., Martí, M., Màrquez, L., Meyers, A., Nivre, J., Padó, S., Štěpánek, J., Straňák, P., Surdeanu, M., Xue, N., and Zhang, Y. (2009). The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*.

Harris, Z. (1951). *Methods in structural linguistics*. Chicago University Press, Chicago.

Hochmann, J., Endress, A. D., and Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115:444–457.

Huang, F. and Yates, A. (2009). Distributional representations for handling sparsity in supervised sequence-labeling. In *ACL*.

Johnson, M. (2007). Why doesn't EM find good HMM POS-taggers? In *Proceedings of the 2007 Joint Conference of EMNLP-CoNLL*, pages 296–305.

Johnson, M., Demuth, K., Frank, M. C., and Jones, B. (2010). Synergies in learning words and their meanings. In *Neural Information Processing Systems, 23*.

Kazama, J. and Torisawa, K. (2007). A new perceptron algorithm for sequence labeling with non-local features. In *Proceedings of the 2007 Joint Conference of EMNLP-CoNLL*, pages 315–324.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99:349–364.

Kingsbury, P. and Palmer, M. (2002). From Treebank to PropBank. In *Proceedings of LREC-2002*, Spain.

Klein, D. and Manning, C. (2004). Corpus-based induction of syntactic structure: Models of dependency and constituency. In *Proceedings of the Association for Computational Linguistics (ACL)*.

Landau, B. and Gleitman, L. (1985). *Language and experience*. Harvard University Press, Cambridge, MA.

Levin, B. and Rappaport-Hovav, M. (2005). *Argument Realization*. Research Surveys in Linguistics Series. Cambridge University Press, Cambridge, UK.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition*. Lawrence Elrbaum Associates, Mahwah, NJ.

Marcus., G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283:77–80.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Màrquez, L., Carreras, X., Litkowski, K., and Stevenson, S. (2008). Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34:145–159.

Meilă, M. (2002). Comparing clusterings. Technical Report 418, University of Washington Statistics Department.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90:91–117.

Mintz, T., Newport, E., and Bever, T. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26:393–424.

Monaghan, P., Chater, N., and Christiansen, M. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, 96:143–182.

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374.

Nappa, R., Wessel, A., McEldoon, K., Gleitman, L., and Trueswell, J. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, 5:203–234.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Parisien, C. and Stevenson, S. (2010). Learning verb alternations in a usage-based bayesian model. In *Proc. of the 32nd annual meeting of the Cognitive Science Society*.

Perfors, A., Tenenbaum, J., and Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, 37.

Pinker, S. (1984). *Language learnability and language development*. Harvard University Press, Cambridge, MA.

Pinker, S. (1989). *Learnability and Cognition*. Cambridge: MIT Press.

Punyakanok, V., Roth, D., and Yih, W. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2).

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285.

Ravi, S. and Knight, K. (2009). Minimized models for unsupervised part-of-speech tagging. In *Proceedings of the Joint Conferenceof the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP)*.

Rispoli, M. (1989). Encounters with japanese verbs: Caregiver sentences and the categorization of transitive and intransitive action verbs. *First Language*, 9:57–80.

Romberg, A. R. and Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1:906–914.

Shi, R., Morgan, J. L., and Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, 25(01):169–201.

Shi, R., Werker, J. F., and Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72(2):B11 – B21.

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

Snedeker, J. and Gleitman, L. (2004). Why it is hard to label our concepts. In Hall and Waxman, editors, *Weaving a Lexicon*. MIT Press, Cambridge, MA.

Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Science*, 102.

Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., and Nivre, J. (2008). The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning*.

Tomasello, M. (2003). *Constructing a language: A Usage-Based theory of language acquisition*. Harvard University Press.

Toutanova, K. and Johnson, M. (2007). A bayesian LDA-based model for semi-supervised part-of-speech tagging. In *Proceedings of NIPS*.

Waterfall, H., Sandbank, B., Onnis, L., and Edelman, S. (2010). An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language*, 37:671–703.

Yang, C. (2011). A statistical test for grammar. In *Proc. of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*.

Yu, C. and Joachims, T. (2009). Learning structural svms with latent variables. In *ICML*.

Yuan, S., Fisher, C., and Snedeker, J. (in press). Counting the nouns: Simple structural cues to verb meaning. *Child Development*.