Putting Meaning into Grammar Learning

Nancy Chang
UC Berkeley, Dept. of Computer Science and
International Computer Science Institute
1947 Center St., Suite 600
Berkeley, CA 94704 USA
nchang@icsi.berkeley.edu

Abstract

This paper proposes a formulation of grammar learning in which meaning plays a fundamental role. We present a computational model that aims to satisfy convergent constraints from cognitive linguistics and crosslinguistic developmental evidence within a statistically driven framework. The target grammar, input data and goal of learning are all designed to allow a tight coupling between language learning and comprehension that drives the acquisition of new constructions. The model is applied to learn lexically specific multi-word constructions from annotated child-directed transcript data.

1 Introduction

What role does meaning play in the acquisition of grammar? Computational approaches to grammar learning have tended to exclude semantic information entirely, or else relegate it to lexical representations. Starting with Gold's (1967) influential early work on language identifiability in the limit and continuing with work in the formalist learnability paradigm, grammar learning has been equated with syntax learning, with the target of learning consisting of relatively abstract structures that govern the combination of symbolic linguistic units. Statistical, corpus-based efforts have likewise restricted their attention to inducing syntactic patterns, though in part due to more practical considerations, such as the lack of large-scale semantically tagged corpora.

But a variety of cognitive, linguistic and developmental considerations suggest that meaning plays a central role in the acquisition of linguistic units at all levels. We start with the proposition that language *use* should drive language learning — that is, the learner's goal is to improve its ability to communicate, via comprehension and production. Cognitive and constructional approaches to grammar assume that the basic unit of linguistic knowledge needed to support language use consists of pairings of form and meaning, or **constructions** (Langacker, 1987; Goldberg, 1995; Fillmore and Kay, 1999). Moreover, by the time children make the leap from single words to complex combinations, they have amassed considerable conceptual knowledge, including familiarity with a wide variety of entities and events and sophisticated pragmatic skills (such as using joint attention to infer communicative intentions (Tomasello, 1995) and subtle lexical distinctions (Bloom, 2000)). The developmental evidence thus suggests that the input to grammar learning may in principle include not just surface strings but also meaningful situation descriptions with rich semantic and pragmatic information.

This paper formalizes the grammar learning problem in line with the observations above, taking seriously the ideas that the target of learning, for both lexical items and larger phrasal and clausal units, is a bipolar structure in which meaning is on par with form, and that meaningful language use drives language learning. The resulting core computational problem can be seen as a restricted type of relational learning. In particular, a key step of the learning task can be cast as learning *relational correspondences*, that is, associations between form relations (typically word order) and meaning relations (typically role-filler bindings). Such correlations are essential for capturing complex multi-unit constructions, both lexically specific constructions and more general grammatical constructions.

The remainder of the paper is structured as follows. Section 2 states the learning task and provides an overview of the model and its assumptions. We then present algorithms for inducing structured mappings, based on either specific input examples or the current set of constructions (Section 3), and describe how these are evaluated using criteria based on minimum description length (Rissanen, 1978). Initial results from applying the learning algorithms to a small corpus of child-directed utterances demonstrate the viability of the approach (Section 4). We conclude with a discussion of the broader implications of this approach for language learning and use.

2 Overview of the learning problem

We begin with an informal description of our learning task, to be formalized below. At all stages of language learning, children are assumed to exploit general cognitive abilities to make sense of the flow of objects and events they experience. To make sense of linguistic events — sounds and gestures used in their environments for communicative purposes — they also draw on specifically linguistic knowledge of how forms map to meanings, i.e., constructions. Comprehension consists of two stages: identifying the constructions involved and how their meanings are related (**analysis**), and matching these constructionally sanctioned meanings to the actual participants and relations present in context (**resolution**). The set of linguistic constructions will typically provide only a *partial* analysis of the utterance in the given context; when this happens, the agent may still draw on general inference to match even a partial analysis to the context.

The goal of construction learning is to acquire a *useful* set of constructions, or **grammar**. This grammar should allow constructional analysis to produce increasingly complete interpretations of utterances in context, thus requiring minimal recourse to general resolution and inference procedures. In the limit the grammar should stabilize, while still being useful for comprehending novel input. A useful grammar should also reflect the statistical properties of the input data, in that more frequent or specific constructions should be learned before more infrequent and more general constructions.

Formally, we define our learning task as follows: Given an initial grammar G and a sequence of training examples consisting of an utterance paired with its context, find the best grammar G' to fit seen data and generalize to new data. The remainder of this section describes the hypothesis space, prior knowledge and input data relevant to the task.

2.1 Hypothesis space: embodied constructions

The space of possible grammars (or sets of constructions) is defined by Embodied Construction Grammar (ECG), a computationally explicit unification-based formalism for capturing insights from the construction grammar and cognitive linguistics literature (Bergen and Chang, in press; Chang et al., 2002). ECG is designed to support the analysis process mentioned above, which determines what constructions and schematic meanings are present in an utterance, resulting in a *semantic specification* (or *semspec*).[1]

We highlight a few relevant aspects of the formalism, exemplified in Figure 1. Each construction has sections labeled **form** and **meaning** listing the entities (or roles) and constraints (type constraints marked with :, filler constraints marked with $\longleftarrow$, and identification (or coindexation) constraints marked with $\longleftrightarrow$) of the respective domains. These two sections, also called the form and meaning **poles**, capture the basic intuition that constructions are form-meaning pairs. A subscripted f or m allows reference to the form or meaning pole of any construction, and the keyword **self** allows self-reference. Thus, the THROW construction simply links a form whose orthography role (or feature) is bound to the string "throw" to a meaning that is constrained to be of type Throw, a separately defined conceptual schema corresponding to throwing events (including roles for a thrower and throwee). (Although not shown in the examples, the formalism also includes a **subcase of** notation for expressing constructional inheritance.)

construction THROW
  form
    self$_f$.orth $\longleftarrow$ "throw"
  meaning
    self$_m$ **:** Throw

construction THROW-TRANSITIVE
  constituents
    t1 **:** REFERRING-EXPRESSION
    t2 **:** THROW-CONSTRUCTION
    t3 **:** REFERRING-EXPRESSION
  form
    t1$_f$ `before` t2$_f$
    t2$_f$ `before` t3$_f$
  meaning
    t2$_m$.thrower $\longleftrightarrow$ t1$_m$
    t2$_m$.throwee $\longleftrightarrow$ t3$_m$

Figure 1: Embodied Construction Grammar representation of the lexical THROW and lexically specific THROW-TRANSITIVE construction (licensing expressions like *You throw the ball*).

Multi-unit constructions such as the THROW-TRANSITIVE construction also list their **constituents**, each of which is itself a form-meaning construction. These multi-unit constructions serve as the target representation for the specific learning task at hand. The key representational insight here is that the form and meaning constraints typi-

[1]ECG is intended to support a simulation-based model of language understanding, with the semspec parameterizing a simulation using active representations (or *embodied schemas*) to produce context-sensitive inferences. See Bergen and Chang (in press) for details.

cally involve *relations* among the form and meaning poles of the constructional constituents. For current purposes we limit the potential form relations to word order, although many other form relations are in principle allowed. In the meaning domain, the primary relation is *identification*, or unification, between two meaning entities. In particular, we will focus on role-filler bindings, in which a role of one constituent is identified with another constituent or with one of its roles. The example construction pairs two word order constraints over its constituents' form poles with two identification constraints over its constituents' meaning poles (these specify the fillers of the thrower and throwee roles of a Throw event, respectively).

Note that both lexical constructions and the multi-unit constructions needed to express grammatical patterns can be seen as graphs of varying complexity. Each domain (form or meaning) can be represented as a subgraph of elements and relations among them. Lexical constructions involve a simple mapping between these two subgraphs, whereas complex constructions with constituents require *structured relational mappings* over the two domains, that is, mappings between form and meaning relations whose arguments are themselves linked by known constructions.

2.2 Prior knowledge

The model makes a number of assumptions based on the child language literature about prior knowledge brought to the task, including conceptual knowledge, lexical knowledge and the language comprehension process described earlier. Figure 2 depicts how these are related in a simple example; each is described in more detail below.

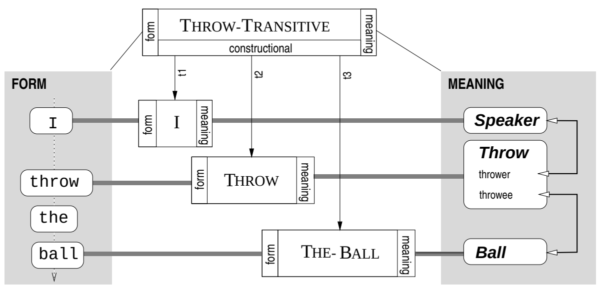

Figure 2: A constructional analysis of *I throw the ball*, with form elements on the left, meaning elements (conceptual schemas) on the right and constructions linking the two domains in the center.

2.2.1 Conceptual knowledge

Conceptual knowledge is represented using an ontology of typed feature structures, or **schemas**. These include schemas for people, objects (e.g. Ball in the figure), locations, and actions familiar to children by the time they enter the two-word stage (typically toward the end of the second year). Actions like the Throw schema referred to in the example THROW construction and in the figure have roles whose fillers are subject to type constraints, reflecting children's knowledge of what kinds of entities can take place in different events.

2.2.2 Lexical constructions

The input to learning includes a set of lexical constructions, represented using the ECG formalism, linking simple forms (i.e. words) to specific conceptual items. Examples of these include the I and BALL constructions in the figure, as well as the THROW construction formally defined in Figure 1. Lexical learning is not the focus of the current work, but a number of previous computational approaches have shown how simple mappings may be acquired from experience (Regier, 1996; Bailey, 1997; Roy and Pentland, 1998).

2.2.3 Construction analyzer

As mentioned earlier, the ECG construction formalism is designed to support processes of language use. In particular, the model makes use of a construction analyzer that identifies the constructions responsible for a given utterance, much like a syntactic parser in a traditional language understanding system identifies which parse rules are responsible. In this case, however, the basic representational unit is a form-meaning pair. The analyzer must therefore also supply a semantic interpretation, called the semspec, indicating which conceptual schemas are involved and how they are related. The analyzer is also required to be robust to input that is not covered by its current grammar, since that situation is the norm during language learning.

Bryant (2003) describes an implemented construction analyzer program that meets these needs. The construction analyzer takes as input a set of ECG constructions (linguistic knowledge), a set of ECG schemas (conceptual knowledge) and an utterance. The analyzer draws on partial parsing techniques previously applied to syntactic parsing (Abney, 1996): utterances not covered by known constructions yield partially filled semspecs, and unknown forms in the input are skipped. As a result, even a small set of simple constructions can provide skeletal interpretations of complex utterances.

Figure 2 gives an iconic representation of the result of analyzing the utterance *I throw the ball* using the THROW-TRANSITIVE and THROW constructions shown earlier, along with some additional lexical constructions (not shown). The analyzer matches each input form with its lexical construction (if available) and corresponding meaning, and then matches the clausal construction by checking the relevant word order relations (implicitly represented by the dotted arrow in the figure) and role bindings (denoted by the double-headed arrows within the meaning domain) asserted on its candidate constituents. Note that at the stage shown, no construction for *the* has yet been learned, resulting in a partial analysis. At an even earlier stage of learning, before the THROW-TRANSITIVE construction is learned, the lexical constructions are matched without resulting in the role-filler bindings on the Throw action schema.

Finally, note that the semspec produced by constructional analysis (right-hand side of the figure) must be matched to the current situational context using a contextual interpretation, or resolution, process. Like other resolution (e.g. reference resolution) procedures, this process relies on category/type constraints and (provisional) identification bindings. The resolution procedure attempts to unify each schema and constraint appearing in the semspec with a type-compatible entity or relation in the context. In the example, the schemas on the right-hand side of the figure should be identified during resolution with particular schema instances available in context (e.g., the Speaker schema should be linked to the specific contextually available discourse speaker, the Ball schema to a particular ball instance, etc.).

2.3 Input data

The input is characterized as a set of **input tokens**, each consisting of an utterance form (a string of known and novel word-forms) paired with a specific communicative context (a set of linked conceptual schemas corresponding to the participants, salient scene and discourse information available in the situation). The learning model receives only positive examples, as in the child learning case. Note, however, that the interpretation a given utterance has in context depends on the current state of linguistic knowledge. Thus the same utterance at different stages may lead to different learning behavior.

The specific training corpus used in learning experiments is a subset of the Sachs corpus of the CHILDES database of parent-child transcripts(Sachs, 1983; MacWhinney, 1991), with additional annotations made by developmental psychologists as part of a study of motion utterances (Dan I. Slobin, p.c.). These annotations indicate semantic and pragmatic features available in the scene. A simple feature structure representation of a sample input token is shown here; boxed numbers indicate that the relevant entities are identified:

$$\begin{bmatrix} \text{Form : } \begin{bmatrix} \text{text : throw the ball} \\ \text{intonation : falling} \end{bmatrix} \\ \text{Participants : Mother } \boxed{0}\text{, Naomi } \boxed{1}\text{, Ball } \boxed{2} \\ \text{Scene : } \begin{bmatrix} \text{Throw} \\ \text{thrower : Naomi } \boxed{1} \\ \text{throwee : Ball } \boxed{2} \end{bmatrix} \\ \text{Discourse : } \begin{bmatrix} \text{speaker : Mother } \boxed{0} \\ \text{addressee : Naomi } \boxed{1} \\ \text{speech act : imperative} \\ \text{activity : play} \\ \text{joint attention : Ball } \boxed{2} \end{bmatrix} \end{bmatrix}$$

Many details have been omitted, and a number of simplifying assumptions have been made. But the rough outline given here nevertheless captures the core computational problem faced by the child learner in acquiring multi-word constructions in a framework putting meaning on par with form.

3 Learning algorithms

We model the learning task as a search through the space of possible grammars, with new constructions incrementally added based on encountered data. As in the child learning situation, the goal of learning is to converge on an optimal set of constructions, i.e., a grammar that is both general enough to encompass significant novel data and specific enough to accurately predict previously seen data.

A suitable overarching computational framework for guiding the search is provided by the minimum description length (MDL) heuristic (Rissanen, 1978), which is used to find the optimal analysis of data in terms of (a) a compact representation of the data (i.e., a grammar); and (b) a compact means of describing the original data in terms of the compressed representation (i.e., constructional analyses using the grammar). The MDL heuristic exploits a tradeoff between competing preferences for smaller grammars (encouraging generalization) and for simpler analyses of the data (encouraging the retention of specific/frequent constructions).

The rest of this section makes the learning framework concrete. Section 3.1 describes several heuristics for moving through the space of grammars (i.e., how to update a grammar with new constructions based on input data), and Section 3.2 describes how to chose among these candidate moves to find optimal points in the search space (i.e., specific MDL criteria for evaluating new grammars). These specifications extend previous methods to accommodate the relational structures of the ECG formalism and the process-based assumptions of the model.

3.1 Updating the grammar

The grammar may be updated in three ways:

hypothesis forming new structured maps to account for mappings present in the input but unexplained by the current grammar;

reorganization exploiting regularities in the set of known constructions (merge two similar constructions into a more general construction, or compose two constructions that cooccur into a larger construction); and

reinforcement incrementing the weight associated with constructions that are successfully used during comprehension.

Hypothesis. The first operation addresses the core computational challenge of learning new structured maps. The key idea here is that the learner is assumed to have access to a partial analysis based on linguistic knowledge, as well as a fuller situation interpretation it can infer from context. Any difference between the two can directly prompt the formation of new constructions that will improve the agent's ability to handle subsequent instances of similar utterances in similar contexts. In particular, certain form and meaning relations that are unmatched by the analysis but present in context may be mapped using the procedure in Figure 3.

Hypothesize construction. Given utterance U in situational context S and current grammar G:

1. Call the construction analysis/resolution processes on (U,S,G) to produce a semspec consisting of form and meaning graphs F and M. Nodes and edges of F and M are marked as matched or unmatched by the analysis.
2. Find $\text{rel}_f(\text{A}_f,\text{B}_f)$, an unmatched edge in F corresponding to an unused form relation over the matched form poles of two constructs A and B.
3. Find $\text{rel}_m(\text{A}_m, \text{B}_m)$, an unmatched edge (or subgraph) in M corresponding to an unused meaning relation (or set of bindings) over the corresponding matched meaning poles A_m and B_m. $\text{rel}_m(\text{A}_m,\text{B}_m)$ is required to be *pseudo-isomorphic* to $\text{rel}_f(\text{A}_f,\text{B}_f)$.
4. Create a new construction C with constituents A and B and form and meaning constraints corresponding to $\text{rel}_f(\text{A}_f,\text{B}_f)$ and $\text{rel}_m(\text{A}_m,\text{B}_m)$, respectively.

Figure 3: Construction hypothesis.

The algorithm creates new constructions mapping form and meaning relations whose arguments are already constructionally mapped. It is best illustrated by example, based on the sample input token shown in Section 2.3 and depicted schematically in Figure 4. Given the utterance "throw the ball" and a grammar including constructions for *throw* and *ball* (but not *the*), the analyzer produces a semspec including a Ball schema and a Throw schema, without indicating any relations between them. The resolution process matches these schemas to the actual context, which includes a particular throwing event in which the addressee (Naomi) is the thrower of a particular ball. The resulting resolved analysis looks like Figure 4 but without the new construction (marked with dashed lines): the two lexical constructions are shown mapping to particular utterance forms and contextual items.

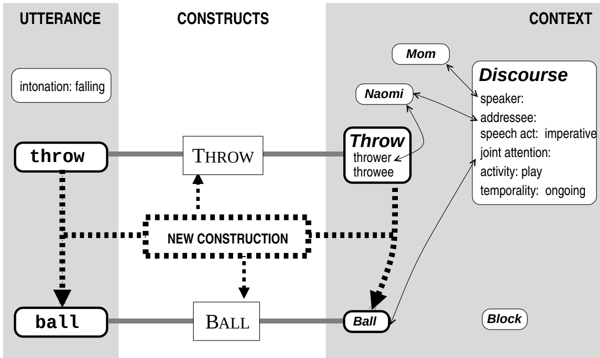

Figure 4: Hypothesizing a relational mapping for the utterance *throw ball*. Heavy solid lines indicate structures matched during analysis; heavy dotted lines indicate the newly hypothesized mapping.

Next, an unmatched form relation (the word order edge between *throw* and *ball*) is found, followed by a corresponding unmatched meaning relation (the binding between the Throw.throwee role and the specific Ball in context); these are shown in the figure using heavy dashed lines. Crucially, these relations meet the condition in step 3 that the relations be *pseudo-isomorphic*. This condition captures three common patterns of relational form-meaning mappings, i.e., ways in which a meaning relation rel_m over A_m and B_m can be correlated with a form relation rel_f over A_f and B_f (e.g., word order); these are illustrated in Figure 5, where we assume a simple form relation:

(a) strictly isomorphic: B_m is a role-filler of A_m (or vice versa) (A_m.r1 $\longleftrightarrow$ B_m)

(b) shared role-filler: A_m and B_m each have a role filled by the same entity (A_m.r1 $\longleftrightarrow$ B_m.r2)

(c) sibling role-fillers: A_m and B_m fill roles of the same schema (Y.r1 $\longleftrightarrow$ A_m, Y.r2 $\longleftrightarrow$ B_m)

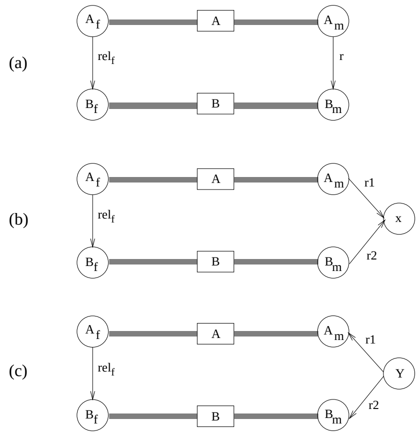

Figure 5: Pseudo-isomorphic relational mappings over constructs A and B: (a) strictly isomorphic; (b) shared role-filler; and (c) sibling role-fillers.

This condition enforces structural similarity between the two relations while recognizing that constructions may involve relations that are not strictly isomorphic. (The example mapping shown in the figure is strictly isomorphic.) The resulting construction is shown formally in Figure 6.

construction THROW-BALL
constituents
t1 : THROW
t2 : BALL
form
$t1_f$ `before` $t2_f$
meaning
$t1_m$.throwee $\longleftrightarrow$ $t2_m$

Figure 6: Example learned construction.

Reorganization. Besides hypothesizing constructions based on new data, the model also allows new constructions to be formed via constructional reorganization, essentially by applying general categorization principles to the current grammar, as described in Figure 7.

For example, the THROW-BALL construction and a similar THROW-BLOCK construction can be merged into a general THROW-OBJECT construction; the resulting subcase constructions each retain the appropriate type constraint. Similarly, a general HUMAN-THROW and THROW-OBJECT construction may occur in many analyses in which they compete for the THROW constituent. Since they have compatible constraints in both form and meaning (in the latter case based on the same conceptual Throw schema), repeated co-occurrence may lead to the formation of a larger construction that includes all three constituents.

Reorganize constructions. Reorganize G to consolidate similar and co-occurring constructions:

- **Merge:** Pairs of constructions with significant shared structure (same number of constituents, minimal ontological distance (i.e., distance in the type ontology) between corresponding constituents, maximal overlap in constraints) may be merged into a new construction containing the shared structure; the original constructions are rewritten as subcases of the new construction along with the non-overlapping information.
- **Compose:** Pairs of constructions that co-occur frequently with compatible constraints (are part of competing analyses using the same constituent, or appear in a constituency relationship) may be composed into one construction.

Figure 7: Construction reorganization.

Reinforcement. Each construction is associated with a weight, which is incremented each time it is used in an analysis that is successfully matched to the context. A successful match covers a majority of the contextually available bindings.

Both hypothesis and reorganization provide means of proposing new constructions; we now specify how proposed constructions are evaluated.

3.2 Evaluating grammar cost

The MDL criteria used in the model is based on the *cost* of the grammar G given the data D:

$$\begin{aligned}
\text{cost}(G|D) &= m \cdot \text{size}(G) + n \cdot \text{cost}(D|G) \\
\text{size}(G) &= \sum_{c \in G} \text{size}(c) \\
\text{size}(c) &= n_c + r_c + \sum_{e \in c} \text{length}(e) \\
\text{cost}(D|G) &= \sum_{d \in D} \text{score}(d) \\
\text{score}(d) &= \sum_{x \in d} (\text{weight}_x + p \cdot \sum_{r \in x} |\text{type}_r|) \\
&\quad + \text{height}_d + \text{semfit}_d
\end{aligned}$$

where m and n are learning parameters that control the relative bias toward model simplicity and data compactness. The size(G) is the sum over the size of each construction c in the grammar (n_c is the number of constituents in c, r_c is the number of constraints in c, and each element reference e in c has a length, measured as slot chain length). The cost (complexity) of the data D given G is the sum of the analysis scores of each input token d using G. This score sums over the constructions

x in the analysis of d, where weight$_x$ reflects relative (in)frequency, $|\text{type}_r|$ denotes the number of ontology items of type r, summed over all the constituents in the analysis and discounted by parameter p. The score also includes terms for the height of the derivation graph and the semantic fit provided by the analyzer as a measure of semantic coherence.

In sum, these criteria favor constructions that are simply described (relative to the available meaning representations and the current set of constructions), frequently useful in analysis, and specific to the data encountered. The MDL criteria thus approximate Bayesian learning, where the minimizing of cost corresponds to maximizing the posterior probability, the structural prior corresponds to the grammar size, and likelihood corresponds to the complexity of the data relative to the grammar.

4 Learning verb islands

The model was applied to the data set described in Section 2.3 to determine whether lexically specific multi-word constructions could be learned using the MDL learning framework described. This task represents an important first step toward general grammatical constructions, and is of cognitive interest, since item-based patterns appear to be learned on independent trajectories (i.e., each verb forms its own "island" of organization (Tomasello, 2003)). We give results for *drop* (n=10 examples), *throw* (n=25), and *fall* (n=50).

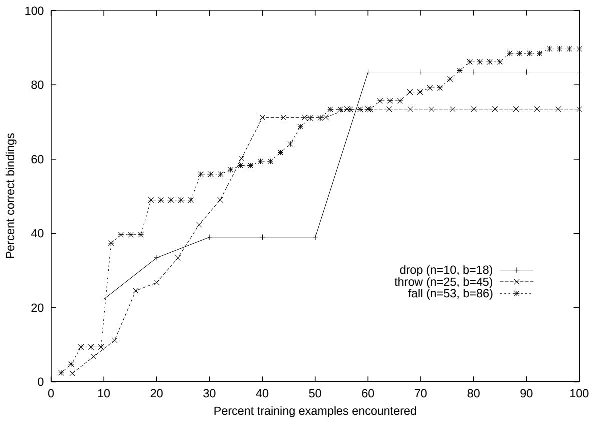

Figure 8: Incrementally improving comprehension for three verb islands.

Given the small corpus sizes, the focus for this experiment is not on the details of the statistical learning framework but instead on a qualitative evaluation of whether learned constructions improve the model's comprehension over time, and how verbs may differ in their learning trajectories. Qualitatively, the model first learned item-specific constructions as expected (e.g. *throw bear*, *throw books*, *you throw*), later in learning generalizing over different event participants (*throw OBJECT*, *PERSON throw*, etc.).

A quantitative measure of comprehension over time, **coverage**, was defined as the percentage of total bindings *b* in the data accounted for at each learning step. This metric indicates how new constructions incrementally improve the model's comprehensive capacity, shown in Figure 8. The *throw* subset, for example, contains 45 bindings to the roles of the Throw schema (thrower, throwee, and goal location). At the start of learning, the model has no combinatorial constructions and can account for none of these. But the model gradually amasses constructions with greater coverage, and by the tenth input token, the model learns new constructions that account for the majority of the bindings in the data.

The learning trajectories do appear distinct: *throw* constructions show a gradual build-up before plateauing, while *fall* has a more fitful climb converging at a higher coverage rate than *throw*. It is interesting to note that the *throw* subset has a much higher percentage of imperative utterances than *fall* (since throwing is pragmatically more likely to be done on command); the learning strategy used in the current model focuses on relational mappings and misses the association of an imperative speech-act with the lack of an expressed agent, providing a possible explanation for the different trajectories.

While further experimentation with larger training sets is needed, the results indicate that the model is able to acquire useful item-based constructions like those learned by children from a small number examples. More importantly, the learned constructions permit a limited degree of generalization that allows for increasingly complete coverage (or comprehension) of new utterances, fulfilling the goal of the learning model. Differences in verb learning lend support to the verb island hypothesis and illustrate how the particular semantic, pragmatic and statistical properties of different verbs can affect their learning course.

5 Discussion and future directions

The work presented here is intended to offer an alternative formulation of the grammar learning problem in which meaning in context plays a pivotal role in the acquisition of grammar. Specifically, meaning is incorporated directly into the target grammar (via the construction representation), the input data (via the context representation) and the evaluation criteria (which is usage-based, i.e. to improve comprehension). To the extent possible, the assumptions made with respect to structures and processes available to a human language learner in this stage are consistent with evidence from across the cognitive spectrum. Though only preliminary conclusions can be made, the model is a concrete computational step toward validating a meaning-oriented approach to grammar learning.

The model draws from a number of computational forerunners from both logical and probabilistic traditions, including Bayesian models of word learning (Bailey, 1997; Stolcke, 1994) for the overall optimization model, and work by Wolff (1982) modeling language acquisition (primarily production rules) using data compression techniques similar to the MDL approach taken here. The use of the results of analysis to hypothesize new mappings can be seen as related to both explanation-based learning (DeJong and Mooney, 1986) and inductive logic programming (Muggleton and Raedt, 1994). The model also has some precedents in the work of Siskind (1997) and Thompson (1998), both of which based learning on the discovery of isomorphic structures in syntactic and semantic representations, though in less linguistically rich formalisms.

In current work we are applying the model to the full corpus of English verbs, as well as crosslinguistic data including Russian case markers and Mandarin directional particles and aspect markers. These experiments will further test the robustness of the model's theoretical assumptions and protect against model overfitting and typological bias. We are also developing alternative means of evaluating the system's progress based on a rudimentary model of production, which would enable it to label scene descriptions using its current grammar and thus facilitate detailed studies of how the system generalizes (and overgeneralizes) to unseen data.

Acknowledgments

We thank members of the ICSI Neural Theory of Language group and two anonymous reviewers.

References

Steven Abney. 1996. Partial parsing via finite-state cascades. In *Workshop on Robust Parsing, 8th European Summer School in Logic, Language and Information*, pages 8–15, Prague, Czech Republic.

David R. Bailey. 1997. *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. Ph.D. thesis, University of California at Berkeley.

Benjamin K. Bergen and Nancy Chang. in press. Simulation-based language understanding in Embodied Construction Grammar. In *Construction Grammar(s): Cognitive and Cross-language dimensions*. John Benjamins.

Paul Bloom. 2000. *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA.

John Bryant. 2003. Constructional analysis. Master's thesis, University of California at Berkeley.

Nancy Chang, Jerome Feldman, Robert Porzel, and Keith Sanders. 2002. Scaling cognitive linguistics: Formalisms for language understanding. In *Proc. 1st International Workshop on Scalable Natural Language Understanding*, Heidelberg, Germany.

G.F. DeJong and R. Mooney. 1986. Explanation-based learning: An alternative view. *Machine Learning*, 1(2):145–176.

Charles Fillmore and Paul Kay. 1999. *Construction grammar*. CSLI, Stanford, CA. To appear.

E.M. Gold. 1967. Language identification in the limit. *Information and Control*, 16:447–474.

Adele E. Goldberg. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar, Vol. 1*. Stanford University Press.

Brian MacWhinney. 1991. *The CHILDES project: Tools for analyzing talk*. Erlbaum, Hillsdale, NJ.

Stephen Muggleton and Luc De Raedt. 1994. Inductive logic programming: Theory and methods. *Journal of Logic Programming*, 19/20:629–679.

Terry Regier. 1996. *The Human Semantic Potential*. MIT Press, Cambridge, MA.

Jorma Rissanen. 1978. Modeling by shortest data description. *Automatica*, 14:465–471.

Deb Roy and Alex Pentland. 1998. Learning audio-visually grounded words from natural input. In *Proc. AAAI workshop, Grounding Word Meaning*.

J. Sachs. 1983. Talking about the there and then: the emergence of displaced reference in parent-child discourse. In K.E. Nelson, editor, *Children's language*, volume 4, pages 1–28. Lawrence Erlbaum Associates.

Jeffrey Mark Siskind, 1997. *A computational study of cross-situational techniques for learning word-to-meaning mappings*, chapter 2. MIT Press.

Andreas Stolcke. 1994. *Bayesian Learning of Probabilistic Language Models*. Ph.D. thesis, Computer Science Division, University of California at Berkeley.

Cynthia A. Thompson. 1998. *Semantic Lexicon Acquisition for Learning Natural Language Interfaces*. Ph.D. thesis, Department of Computer Sciences, University of Texas at Austin, December.

Michael Tomasello. 1995. Joint attention as social cognition. In C.D. Moore P., editor, *Joint attention: Its origins and role in development*. Lawrence Erlbaum Associates, Hillsdale, NJ. Educ/Psych BF720.A85.J65 1995.

Michael Tomasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, MA.

J. Gerard Wolff. 1982. Language acquisition, data compression and generalization. *Language & Communication*, 2(1):57–89.