Chapter 9

Conclusions

Beginnings are always messy.
— *John Galsworthy*

The ideas explored in the foregoing are, of course, only a beginning: the general framework and specific model proposed here constitute but the first steps toward a satisfactory theory of the structure, acquisition and use of language. In this concluding chapter I take stock of where we are, how we got here and how we might proceed.

9.1 Recapitulation

In this work I have endeavored to draw a detailed picture of how meaning and usage in context facilitate the transition from single words to complex utterances. The learning model at the heart of this approach is consistent with a host of evidence from across the cognitive spectrum; it provides a formally precise realization of the foundational principles set out in Chapter 1:

- The target learning representation is a *construction*: the ECG formalism defined in Chapter 3 can express complex relational mappings over the domains of form and meaning.

- Meaning representations have an *embodied* basis: language is assumed to associate forms with meaning schemas that parameterize more detailed embodied structures.

- Learning depends on *usage*: learning is tightly integrated with the language understanding processes described in Chapter 4.

These constraints, along with the considerations reviewed in Chapter 2, motivate both the class of language learning problems defined in Chapter 5 and the class of optimization-based solutions proposed. In particular, I have characterized *learning relational constructions* as an instance of the more general case of language learning, and designed the learning operations of Chapter 6 and the simplicity-based evaluation criteria of Chapter 7 accordingly. Together these chapters instantiate a solution that has been experimentally validated for a subset of early motion-oriented English constructions, as described in Chapter 8. While much further investigation is called for, I hope to have demonstrated the soundness of the underlying premise and provided at least an existence proof of how, given an appropriate formulation, the mechanisms of language acquisition can be rendered less mysterious than they might seem at first blush.

The main proposals embodied by the model can be summarized as follows:

- The acquisition of grammatical constructions depends on many of the same cognitive abilities and pressures as the acquisition of lexical constructions: the tendency to acquire concrete constructions first, in a bottom-up, piecemeal fashion; the tendency to group and generalize similar constructions; and the importance of the general communicative setting in which learning occurs.

- What sets grammatical constructions apart from (simple) lexical constructions is their *relational* nature, as manifested by both constituent structure and relational constraints. The greater complexity of these relational structures has implications for the kinds of formalisms appropriate for representing grammatical knowledge, as well as the strategies needed for positing and evaluating new candidate constructions.

- Learning takes place relative to the learner's current best analysis, employing all background and situational knowledge available — crucially including previously learned constructions. The learner acquires new constructions to bridge the gap between linguistically analyzed and contextually inferred meaning.

- Learning is characterized as an incremental optimization process, in which both the search for new constructions and the evaluation of possible grammars depend on domain-specific and -general heuristics. Minimum description length provides a reasonable simplicity-based background framework for guiding this process, while the immediate needs and results of context-dependent language comprehension provide the raw materials for learning.

Both the high-level structure of the model and the concrete claims made with respect to the particular subproblem at hand should most properly be considered starting points for refinement and experimentation. From a methodological perspective, the key contribution of the model is that it provides a means of stating such assumptions and claims explicitly enough to support more detailed investigations. The model also provides suggestive evidence relevant to more general questions in the study of language acquisition.

In this concluding chapter, I direct my attention outward to discuss the model in its broader scientific context. I survey the most closely related ideas and predecessors of this work, many of which have been mentioned already, and examine some salient distinctions within the relatively small set of comparable approaches. I then consider some paths forward, in particular drawing attention to ways in which the simplifying assumptions made in this model can be relaxed. In fact, there are ample connections between the work presented here and other well-developed research areas; these hold significant potential for illuminating future directions of research. Finally, I consider some more general implications of the model and its potential descendents.

9.2 Nearest neighbors

Along with many other language-oriented researchers, I have elected to take a maximally inclusive approach to language acquisition and understanding in this work, in the hope that the combined constraints of multiple fields will succeed where isolated perspectives have heretofore failed. To a large extent I have focused on the linguistic and developmental influences on the model, and I hope the imprint of construction grammarians and developmental experts remains apparent through all the computational formalities. In particular, the larger family of cognitively and constructionally motivated approaches to language representation and acquisition has directly inspired many of the choices made in the implementation of this model.

In part this focus is due to the relative paucity of computational work subject to the kinds of constraints taken on here: the combination of embodiment, constructional and developmental considerations has not historically found much traction in computational settings. But while there are relatively few efforts that are directly comparable to the current work, the model does have several computational forerunners and nearest neighbors from both logical and probabilistic traditions.

9.2.1 Statistical forerunners and contemporaries

The most direct line of ancestry can be traced in the realm of optimization-based learning approaches, though these vary in how — or whether — they are motivated by cognitive considerations. As mentioned in Chapter 2, Horning's (1969) early work in grammatical inference can be seen as anticipating the probabilistic revolution throughout artificial intelligence. The current model, though far removed in its explicit cognitive and linguistic goals from the kinds of grammar addressed by Horning, is its direct descendent, by way of the model merging approaches of Stolcke (1994) and Bailey (1997). Wolff (1982) also describes several models that take simplicity-based optimization as the key to language acquisition. Like the current model, Wolff's model exploits a number of operations motivated by compression, including some analogous to the generalization and relational mapping operations described here.

Several more recent lines of work address aspects of language learning and use from a Bayesian and information-theoretic perspective consonant with that taken here. (See Pereira (2000) for an overview of information-theoretic approaches to formal linguistics, placed in historical context.) As noted in Chapter 7, Goldsmith (2002) employs an MDL criterion to learn to perform unsupervised morphological segmentation of a variety of languages. While the model is not applied to naturalistic child-directed data, Goldsmith makes a broader case for probabilistic models and optimization-based strategies for evaluating linguistic theories relative to corpus data.[1]

In a more explicitly cognitive vein, Chater (2004) and Chater & Vitányi (2003) similarly argue for an MDL-based principle of simplicity as a general cognitive principle, with particular applications to the child language problems of overgeneralization and learnability (Chater & Vitányi 2007; Chater & Manning 2006). Perfors *et al.* (2006) and Perfors (2008) likewise address the poverty of the stimulus problem, showing how an ideal learner might employ principles of Bayesian model selection to choose among candidate grammars of varying size and complexity, based on child-directed input data. Their results suggest that a preference for hierarchical grammars (in this case, context-free grammars over regular grammars) need not be innately encoded but can rather be inferred from data. Finally, S. Edelman and colleagues (Edelman 2004; Edelman *et al.* 2004) have developed several models of language acquisition that, like the current model, attempt to bring cognitive, biological and statistical ideas together in a framework that is compatible in some of its assumptions with construction-based approaches to grammar.

[1]See also the compression-inspired segmentation models developed by de Marcken (1995); Brent & Cartwright (1996); and Brent (1999); and Clark's (2001) work on unsupervised syntax learning using distributional and MDL-based clustering. Klein & Manning's (2004) models of syntax learning, also based on statistically driven clustering, is notable for its explicit recognition of the value of underlying constituent and dependency structure.

Most of the approaches above explicitly eschew semantic structure, defining their evaluation metrics in terms of structural and form-based properties. In contrast, Alishahi & Stevenson (2008; Alishahi 2008) describe a computational model for learning verbal argument structure patterns, approximating constructions as probabilistic associations between syntactic and semantic patterns. It thus comes closer to addressing the form-meaning mapping problem discussed here, though without explicit constructions of the kind discussed in the construction grammar literature. These works represent a welcome move toward applying well-established Bayesian machine learning techniques to domains that more adequately capture the human language learning scenario.

9.2.2 Logical approaches to language acquisition

Interestingly, some of the most closely related work goes back several decades. Selfridge (1986) describes a system that is remarkably similar in spirit to the current effort, despite major differences in its particulars. Selfridge's system is explicitly intended to model the first five years of a child's linguistic development, where language understanding (in particular, partial understanding), the context of social interaction, and the interaction of comprehension, production and context lead to new grammar rules and lexical items in a simulated conversational environment. Of course, neither the linguistic assumptions nor the simplicity of the underlying formal representation scale well to phenomena beyond simple slot-filler relationships, and the system does not exhibit any significant ability to generalize. Still, it remains a rare and early attempt to harness computational precision in service of cognitive goals.

Several previous models cast language learning as a mapping problem, *i.e.*, one of finding the appropriate correspondences between linguistic forms and meaning representations. The model has some precedents in the work of Siskind (1997) and Thompson (1998), in which relational semantic representations are learned for individual lexical items (Thompson 1998; Thompson & Mooney 1999; Siskind 1997), based on the discovery of isomorphic structures in syntactic and semantic representations. Both of these learn lexical mappings, where sentential semantics are assumed to be compositional from their component words. The mappings themselves are thus not structured, but simple, in the terms used here. Gentner's (1983) work on the importance of recognizing and mapping relational structure in many domains of reasoning, learning and knowledge representation (Gentner & Markman 1997; Markman 1999; Gentner & Namy 2006) and the related Structure Mapping Engine (Falkenhainer *et al.* 1989) resonate more closely with the concerns explored here, particularly the need for finding structural alignment across domains.

The interaction with an analysis procedure to explain data by hypothesizing new mappings can be seen as a variant of explanation-based learning and generalization (DeJong & Mooney 1986); the resulting structured relational mappings are inferred in ways reminiscent of those in inductive logic programming (Muggleton & Raedt 1994).

9.2.3 Grounded language learning and use

Another relevant stream comes from computational models of grounded language learning. These models take a bottom-up approach that emphasizes the situated nature of language learning, exposing robotic and simulated agents to sensorimotor input accompanied by linguistic input in a dynamic environment (Roy 1999; Oates *et al.* 1999; Steels & Kaplan 1998; Steels 1997; Siskind 2001; Cangelosi 2006; Cangelosi 2005; Dominey & Boucher 2005). Work in this area has focused on grounding lexical acquisition in concrete physical domains, with learning scenarios that take the notion of embodiment more literally than feasible for the current model and directly address the symbol grounding problem. Roy (2005), for example, offers a comprehensive theory of language grounding that links a wide variety of signs (in the semiotic sense) to situationally grounded schemas for action and perception. While these signs are not specifically aimed at the representational level addressed here and do not correspond directly to constructions as defined here, the larger semiotically inspired framework provides a useful perspective for reconciling symbolic and analog representations.

The work of Steels and his colleagues is especially compatible with the background assumptions taken here. In a series of experiments inspired by Wittgenstein's (1958) notion of a *language game*, Steels has shown how grounded social interaction among robotic and simulated agents leads to the development of stable communicative mappings that are the beginning of language, including early syntax (Steels 2006; Steels 2000; Steels 1998; Steels & Vogt 1997). Related work in developing Fluid Construction Grammar (De Beule & Steels 2005; Steels *et al.* 2005) is also one of the only other explicitly computational attempts to formalize constructional approaches to grammar, with specific attention to the need for and emergence of compositional, hierarchical structure (De Beuele 2008; De Beule & Bergen 2006). This line of research is most directly motivated by interest in the population dynamics driving the evolution of linguistic communication, and therefore not tuned to the particular constraints of the human learner *per se*. Nonetheless, the overall emphasis on emergent, socially sanctioned mappings of form and meaning is consistent with the goals of the current model, which focuses instead on a single, unequal dyad of agents, one of whom has a fully developed language capacity that is transferred to the other over the course of many interactions. These two complementary views of language learning offer much promise for fruitful integration.

9.3 Onward and upward

Our current position affords many paths forward. Both the class of problems addressed and the class of solutions set forth offer a wealth of opportunities for exploration. This section considers some of the model's limitations, many signalled by the disclaimers made in earlier chapters, and discusses how and to what extent the basic underpinnings of the model may be able to withstand the potential barrage of multidisciplinary objections and concerns. But I optimistically view most current limitations as opportunities for future extensions, rather than intrinsic shortcomings of the model. Some of the most promising directions are outlined here.

9.3.1 Scaling further linguistic heights

Can the construction representation scale to complex linguistic phenomena? What does this model have to do with grammar? What about argument structure, alternations, and the rest of the usual linguistic suspects?

The relations between linguistic form and meaning can be much more complicated than those encountered in the domain space of early constructions, which provides a natural buffer against complexity. But while the examples learned by the model are unlikely to cause much of a stir in an introductory syntax class, the formalism is designed to be not just for kids. In this work I have cast a fairly tight net over a limited set of phenomena that illustrate a key representational challenge, the presence of structured relational mappings between form and meaning. The model can not only acquire such concrete relational mappings from contextually rich input data, but also generalize beyond these to include partially abstract structures with semantically motivated variable components, ranging from item-specific constructions to more general argument structure constructions.

In more recent work, ECG has been extended to handle a wider range of syntactic phenomena involving predicate-argument structure (Feldman *et al.* To appear), and Dodge (In Preparation) presents a detailed analysis of how ECG and the simulation-based framework can cope with classic problems in argument structure alternations. As mentioned in Chapter 5, a probabilistic version of the analyzer process is the subject of Bryant's (2008) dissertation research. Mok (2008a) has also extended many aspects of the current model to accommodate the challenges of more heavily discourse-dependent languages like Mandarin, which feature omitted arguments. More generally, both the formalism and the learning model have been designed with polyglot potential in mind:

though the examples discussed here have been nearly exclusively in English, small-scale studies have demonstrated the formalism's representational and algorithmic potential for accommodating crosslinguistic phenomena, including morphological markers of case and verbal inflection (in, *e.g.*, Russian, Georgian, Hebrew, Spanish and German). Besides shedding light on many outstanding issues in learning theory, these studies serve as a minimal safeguard against typological biases, as well as a foundation for larger-scale experimental validation.

But an embarrassment of riches remains with respect to future work. Perhaps most glaring is the need for a model of language production to complement the models of comprehension and learning that have been the locus of this research. Such a model would not only provide further possibilities for usage-based learning operations, but it would also allow a much more natural means of testing the learning model and replicating experiments from the developmental literature. More broadly, many challenges remain in integrating the specifically linguistic processes of language use with the larger simulation-based framework for modeling inference and belief state, including beliefs and goals in both concrete and metaphorical domains (Narayanan 1997a; Narayanan 1997b; Lakoff & Johnson 1980). All of these could ultimately ground the acquisition of metaphor and metaphorical language based on (and in conjunction with) conceptual and constructional acquisition in more directly embodied domains (Johnson 1999; Grady 1997). As discussed by Chang *et al.* (2002a) and Mok *et al.* (2004), the representational devices of ECG and the simulation-based framework can be extended to represent much more complex linguistic relations, including a variety of mental space and conceptual blending phenomena (Fauconnier 1985; Fauconnier 1987; Fauconnier & Turner 2003). Although further investigation is needed to model such phenomena in detail, the formal framework established here for both learning and understanding language has thus far proven a stable and robust foundation upon which to build.

9.3.2 Modeling developmental trajectories

What can the model tell us about the child learner? How does the artificial learning situation relate to the problem faced by children? To what extent can the assumptions about input data, processing capacities and learning strategies be relaxed?

Computational modeling by its nature forces abstractions and simplifications of the phenomena under study; the choices made for a particular model reflect an implicit claim about which aspects of the problem are most relevant, given the scientific priorities of the modeler, and which can be safely elided. In this case, my priority has been to build a general architecture for language learning that places meaning, context and usage on a par with formal and structural properties of language. This architecture is broadly consistent with the developmental findings reviewed in Chapter 2, in particular the functionalist and emergentist approaches to language acquisition. Fundamentally, however, it is intended to be inclusive with respect to potential inputs to learning, and agnostic about which of these will prove relevant for a particular language, learner or phenomenon. It is similar in this respect to Hirsh-Pasek & Golinkoff's (1996) Emergentist Coalition Model, and it borrows much in spirit (if not implementation) from various proposals of MacWhinney and his colleagues (MacWhinney 2004; MacWhinney 1987; Bates & MacWhinney 1987). The learner of the current model is seen as essentially opportunistic, availing itself of whatever strategies and sources of information most effectively improve its ability to make sense of its environment. These strategies can in theory encompass many flavors of bootstrapping, not just syntactic or semantic but also contextual and pragmatic; both statistical and more deductive, relational styles of learning; and usage in all of the senses distinguished in Section 2.2.3, including both the processes and functions of use, as applied to a single utterance or aggregated over many utterances.

As discussed in Chapters 6 and 7, both the search for learning operations and the evaluation of those operations have direct analogues in some prominent proposals in the developmental literature. While this is not entirely a coincidence, it is encouraging to note that the formal framework developed here has independent motivation from both the statistical and information-theoretic worlds. Although the standard formulations of optimization-based learning, either probabilistic or minimum description length, require some adaptation to eliminate unreasonable assumptions and recognize the special needs of the human learner, the underlying bias toward simplicity appears well-founded in the child domain. As noted earlier, Clark's (2003) principles of simplicity and transparency are particularly clear examples of this convergence, and the specific learning operations proposed by Slobin (1985) also resonate with both the search and evaluation strategies of the model. Indeed, it seems plausible that with relatively simple extensions, many if not all of the proposed operations could find a home within the uniform framework provided by the model. The many findings and usage-based proposals of Tomasello (2003) and his colleagues have also provided much foundational inspiration for the current model.

Of course, the case studies presented here barely scratch the surface of language learning phenomena to be studied, in terms of both theoretical issues to address and empirical findings to explain. I have proposed a few basic mechanisms by which a learner can form new constructions, along with an evaluation metric that incorporates notions of simplicity and usefulness to choose among them; both of these can and should be extended to exploit other domain-general and -specific heuristics. Other operations could, for example, make use of a greater variety of form and meaning relations; finer-grained segmentation of phonological, intonational and morphological information; simulation as a source of inferred (but not situationally perceived) meanings (see Section 9.3.3 below); and production-based operations that allow linguistic exploration to be rewarded or punished via reinforcement.

Regardless of the theoretical ground covered, however, matching the model more closely to human performance capacities and limitations would require significantly more data than currently available, in two senses: (1) input training data for the model, annotated as needed with appropriate contextual information; and (2) experimental results illuminating the mechanisms of human language learning and use, especially as they relate to the framework proposed here. The collection and annotation of data appropriate for a usage-based model poses a non-trivial challenge, whether done (semi-)automatically or manually; much work remains to establish methods and standards that are both crosslinguistically sound and flexible enough to accommodate a range of theoretical assumptions. Relevant empirical data on child and adult acquisition (and use) are much more widespread, and some of the psychological findings discussed in Section 2.1.3 and Section 2.2.3 provide potential grounds for experimentation and replication. Recent interest in statistical learning has been especially helpful for encouraging more systematic study of the nature of the input data, as well as the development of experimental paradigms that allow controlled manipulation of the input (Wonnacott *et al.* 2008; Hudson Kam & Newport 2005; Goldberg *et al.* 2004; Gómez & Gerken 2002; Gómez & Gerken 2000). All of these provide promising opportunities to determine the degree to which the model can exhibit behaviors attested in human learners.

Several avenues of research could relax the assumptions of the model to address a broader set of phenomena. The model does not require the strict separation of lexical and constructional learning, but a fuller integration of these processes, along with the acquisition of complex morphosyntactic constructions, is certainly required. Various lexical learning models (such as those mentioned in Section 9.2.3, and the previous NTL models of Bailey (1997) and Regier (1996) incorporate more directly embodied, situated representations; the representations and techniques used to acquire lexical mappings to these richer, more grounded domains could be better integrated with the acquisition of complex relational constructions addressed here. Relatedly, concept learning itself could be interleaved with language learning, where the usage-based learning techniques proposed here could directly prompt structural alignment and generalization processes that lead to the formation of new linguistically and statistically motivated conceptual categories. More explicit modeling of crosslinguistic variation and invariants should also be undertaken to investigate how easily the usage operations and evaluation criteria adapt to different learning circumstances. These might shed light on the universality (or specificity) of various language learning operations and allow the model to engage more directly with Whorfian issues around the interplay between language and concept.

9.3.3 Situational grounding

Can the model scale to real-world input? To what extent does the input representation simplify the learning problem? How does simulation fit into the model?

This work has deliberately sidestepped some of the difficulties of learning in the real world: the schema-based representations used as input are a far cry from the sounds and sensations of a continuous environment experienced by human (or robotic) learners. Moreover, humans (and other animals) must draw on a wide variety of non-linguistic cues to infer goals, solve problems and imagine consequences. It appears likely that language learning, and language use more generally, is AI-complete, so any attempt to simulate human language learning with more fidelity will require more integrated models of all of these aspects of human behavior and cognition.

It seems reasonable, however, to assume that these challenges — *e.g.*, scene parsing, event and plan recognition, the inference of agent goals and intentions — are theoretically separable from the learning problem addressed in the current work. After all, the separate, parallel development of models that target different aspects of an enormously complex phenomenon is the bread and butter of scientific progress. In this case, there are both developmental and representational reasons to endow our learner with some preprocessing abilities. Young children's ability to infer intentions and achieve goals seems to develop in large part before they have progressed very far along the path to language, and certainly well before our primary stage of interest in this work. Further, no amount of raw sensory data will support the acquisition of complex relational constructions if the appropriate representations are not available.

That said, it would be desirable for the current model to scale up to more naturalistic input data. The work cited in Section 9.2.3, especially that of Steels (2006) and Roy (2005), demonstrate that raw sensory input can be processed as a preliminary step within a language learning system, potentially producing relational predicates similar to the input data assumed here. An alternate direction of development would be toward a more fluid representation of the ongoing flow of events and utterances. As noted earlier, Mok (2008a) has extended the current approach to encompass data consisting of multiple utterances within a contiguous learning episode, embedded within a structured context model tracking extended situational and discourse history (Chang & Mok 1998). That is, the learner must determine how utterances map to referent objects and events. Both kinds of scaling impose additional demands on the learner by introducing more referential ambiguity and indeterminacy of speaker intent. Note that in some sense such modifications simply increase the potential for noisy input (*i.e.*, the learner has more candidates to choose from when mapping an utterance to its referent events and objects), without affecting the model's basic representational assumptions. They might also, however, have considerable practical advantages, since they would reduce the need for manual annotation efforts that incur great labor while introducing nagging worries about unwarranted input presumptions.

Although the learning model does not make direct use of the simulation engine, the idea of dynamic simulation as a means of generating inferences and interpreting language remains a crucial one for our purposes. Just as reference resolution provides a fallback measure for dealing with uninterpretable sentences, simulation could function as yet another process that relieves the encoding burden on the learner. That is, the learner is free to learn something as minimal as the schemas and bindings among them precisely because it is reasonable to assume that the processes of resolution (and simulation) can supply the rich dynamic details that apply in context. A straightforward extension of the model would allow learning to use the results of not just analysis and resolution but also simulation to guide the search for new operations. New linguistic constructions would be biased toward connections that are either not directly evident from resolution and simulation, or else frequent, useful and salient enough to justify the cost of encoding them.

9.3.4 Biological plausibility

What's neural about this? What biological structures and mechanisms does the model map to? How does it relate to connectionist approaches?

The current work focuses on the computational level and how it can felicitously capture phenomena at the cognitive and linguistic level. In keeping with the layered methodology of the NTL project (as described in Chapter 1 and, in much greater detail, by Feldman (2006)), the representational toolkit used by the learning model is intended to derive from biologically plausible mechanisms, and in particular those that have plausible implementations based on structured connectionist models (Shastri *et al.* 1999). These share many basic assumptions with other connectionist approaches (Rumelhart *et al.* 1986; Elman *et al.* 1996), but are distinguished by their emphasis on the highly structured nature of neural representation. The computational formalisms employed to represent relational constructions, albeit more complex than those used in other work (*e.g.*, Bailey 1997), nonetheless can in theory exploit the same reductions to the structured connectionist level. Specifically, feature-based conceptual representations capturing relational role-filler bindings can be approximated using functional clusters of units called *triangle nodes* (Shastri 1988; Shastri & Ajjanagadde 1993), and Bayesian model merging has a connectionist realization in *recruitment learning* and other *vicinal* algorithms (Shastri 2001); Valiant (1984) also argues for the biological plausibility of vicinal algorithms based on the theoretical computational and storage properties of the brain.

The broader principles driving the model are also inspired by and compatible with many biologically motivated proposals in the literature. In general, the concern with embodiment as the basis of meaning and the view of linguistic constructions as linking cross-domain (and neurally grounded) representations is a much more natural fit with biologically minded researchers than formalist approaches to language; G. Edelman 2007, for example, characterizes cognitive approaches to semantics as a return to biology that better comports with the developing view of how concepts and categories are grounded in the body and brain.

More explicitly, the Simulation Hypothesis is motivated in part by an exciting set of results over the last decade on *mirror neurons* (Gallese *et al.* 1996; Rizzolatti *et al.* 1996), neural systems found in primates and humans that are active in both the recognition and execution of highly specialized, goal-based actions, such as grasping or lip-smacking. These mirror systems have been found to be active during sentence comprehension (Buccino *et al.* 2001; Tettamanti *et al.* 2005), and Gallese & Lakoff (2005) argue that they serve as the basis for embodied concepts and embodied simulation. The mirror system's putative role in imitation has also fueled much speculation about how it may have spurred language evolution. Deacon (1997) and Tomasello (1999) both focus on the emergence of the sign, while the Mirror System Hypothesis of Arbib & Rizzolatti (1997; Rizzolatti & Arbib 1998; Arbib 2006) proposes a progression from imitative behaviors in early hominids to communicative manual and vocal gestures, and finally to combinatory language in humans, all grounded by the mirror system. Interestingly, a key stage of the proposed trajectory requires the ability to decompose actions into their component parts for later recombination in novel contexts. This focus on composition and the basis of generalized predicate-argument relations — in both action, and eventually, language — is precisely the functional mechanism targeted in this work as the key challenge children face in moving from the single-word stage to combinatorial grammar.

9.3.5 Natural and artificial intelligence

Can the model scale up from artificial examples to situations with vast amounts of data? Does it have applications separate from its cognitive motivations? How does it relate to statistical machine learning?

Models of language learning are, like other areas addressed by research in artificial intelligence, subject to the tension between wishing to deploy to the utmost our statistical and computational resources to address the relevant abstract problem and recognizing that constraints on actual (embodied) cognitive systems may change the nature of the problem in fundamental ways. Fortunately, these approaches are not mutually exclusive, and there is much ground for refining the current model to better exploit the myriad tools of statistical learning theory (as exemplified by some of the related work discussed in Section 9.2.1), reinforcement learning and relational learning.

Some of the kinds of inductive bias adopted in this task — *e.g.*, the particular linguistic and ontological representations used, and the structural priors based on minimum description length — are reminiscent of much previous work. But the approach described here emphasizes the influence of background world knowledge, the interim results of (online) learning and the overarching communicative goals and processes. The available input data provides only indirect evidence of the target of learning and is thus subject to structural analysis using previously learned constructions (as well as general knowledge). Importantly, the analysis is closely linked to the task for which the target of learning is used — *i.e.*, language comprehension — giving rise to an additional performance-based inductive bias.

Although intended primarily as a model of child language learning, the formalisms and algorithms used here may be applicable to less cognitively oriented domains as well. The formalization of ideas from cognitive linguistics, for example, addresses a much wider range of natural language phenomena than most approaches used in computational linguistics, and may be of use for semantically demanding tasks like question answering (Sinha 2008) and machine translation. Moreover, the learning techniques here could potentially be applied to semantically tagged data and lexical resources that become available, exploiting the learning-analysis cycle's ability to induce complex grammars in bootstrap fashion from simpler data.

9.4 Implications

The viability of the proposed approach to learning grammatical constructions has several potential implications.

9.4.1 Poverty and opulence revisited

Any model of the acquisition of grammar — even in the nascent form encountered here — necessarily treads into the dangerous territory of innateness and modularity. As discussed in Section 2.1.2, different formulations of the problem make radically different assumptions and conclusions. The foundational assumptions motivating the model proposed here align it squarely with the emergentist, interactionist view of acquisition. Indeed, it is explicitly designed to investigate how and whether such theories can be formally realized, in that it makes minimal assumptions about specifically linguistic biases while exploiting cognitively motivated structures and processes. The inclusion of meaning in every aspect of the problem leads to an approach that differs markedly from what has become the conventional wisdom in the field. While this approach introduces new representational challenges, it also allows the learner to exploit all available information and thereby face an easier — or in any case different — task, characterized not by impoverished input but by a wealth of semantic and pragmatic constraints.

To the extent that it makes reasonable assumptions about the experience brought to the task by children entering the two-word stage, it suggests that domain-general learning principles can lead to the acquisition of multi-unit expressions, without domain-specific biases of the kind typically proposed by nativist theories. It also demonstrates how the paradox set up by formal learnability theory can be neatly sidestepped by relaxing the assumptions in accord with linguistic and developmental evidence. In bolstering the view that children learn constructions, it weakens the argument for otherwise unmotivated assumptions about innately endowed syntactic principles.

It is worth observing that the model does not inherently preclude genetically encoded biases — in fact, the appeal to embodiment explicitly relies on neural and evolutionary forces that are nothing if not innately endowed. Moreover, our preoccupation with the building of constituent structure mirrors the rarefied status of recursion in the Chomskyan paradigm as the defining characteristic of the human capacity for language (Hauser *et al.* 2002), though such a bias can be readily interpreted as domain-general. Finally, nothing in the proposed model is intended to deny the many genuinely vexing phenomena uncovered by research in the nativist and syntacto-centric paradigm. It seems possible, however, that some of these puzzles might be more productively tackled within a framework that more closely approximates the child's cognitive capacities. The resolution of such issues remains an empirical matter.

9.4.2 Against spurious dichotomies

The themes just sounded — against the longstanding dichotomy between nature and nurture, and between domain-general and -specific mechanisms in language learning — are part of a larger motif arguing against polarizing tendencies in the study of language. While useful distinctions can and should be made, the recurring theme suggests that, where language is concerned, theoretical poles might be more usefully viewed as working in concert than in opposition.

The construction grammar ethos is itself explicit about the harmonious integration of form and meaning, as expressed in each constructional mapping — at all levels of size, granularity and abstraction, and in both central and peripheral kinds of constructions. It also places lexicon and grammar on a continuum rather than a divide. The ECG variant further blurs many categorical divisions within the meaning domain: linguistic meaning, embodied semantic schemas, pragmatic inference based on situational and discourse context are all connected ultimately through simulation, and the interface resulting from the dual status of embodied schemas — as specifying both linguistic meaning and parameters for simulations — further underscores the essential connection between static structures and dynamic processes.

This tight connection is apparent, too, in the integrated theories of language structure, use and acquisition: constructions serve as both the supporting infrastructure for language understanding and the target of language learning, and mechanisms of learning exploit (partial) language understanding to improve those same processes by improving the current set of constructions. From a psychological perspective, the information-theoretic view of learning likewise finds some common currency for storage and processing: both rote memorization and structure-based generalization are driven by the need to encode experience compactly, balanced against the predictive value of encoding it more completely. Computationally, learning strategies inspired by symbolic and logical insights are deployed as part of a larger probabilistically motivated framework.

The apparent fluidity of these various traditional dichotomies may not be entirely surprising, under the assumption that all of these structures and processes are ultimately realized at the neural level. It remains to be seen how and whether these distinctions will persist in theories and models closer to the neural and structured connectionist levels of explanation.

9.4.3 Toward a cognitive science of language learning

Two paradoxes are better than one; they may even suggest a solution.
— *Edward Teller*

Language learning may in a literal sense be child's play, but nothing about the scientific problems it raises is easy. The ideas proffered here are intended to bridge a few of the gaps in our collective (and as yet partial) understanding of the problem. Given the number of simplifying assumptions and abstractions necessary at this stage, it may be unlikely that the proposed solution will be wholly satisfactory by the lights of any particular disciplinary perspective. Still, I hope to have achieved the broader goal of demonstrating how the various overlapping magisteria can be brought together to nudge longstanding debates on the nature of the problem a bit further along the path from quasi-religious war to rigorous science.

Indeed, I hope to have suggested that a more ecumenical stance might ultimately be more economical as well — that, paradoxically, embracing the unfamiliar habits and limitations of one's disciplinary neighbors may reveal a graceful middle way forward not available from more restricted vantage points. I have shown, on the one hand, how attention to cognitive constraints on human language structure, use and acquisition can have far-reaching consequences for how we define the problem space and what solutions are adequate; and on the other, how submission to formal and computational methods can endow our investigations with some of the order and precision necessary for coping with the staggering complexity of the phenomena involved.

It is one thing to build a model or theory that heeds the concerns of multiple constituencies; it is quite another to establish common ground for serious ongoing interdisciplinary engagement. To the latter end, this work is intended to be not just about constructing a grammar, formalism or model; rather, it is meant as a starting point for a more informed, inclusive and productive dialogue, in which assumptions can be challenged and changed as we learn more about our joint endeavor. However contentious this conversation will assuredly be, I am certain that the resulting holistic whole will be greater than the sum of its constituent parts.