Chapter 1

Beyond single words

How do children make the leap from single words to complex combinations? The simple act of putting one meaningful unit after another is a defining characteristic of linguistic competence. It is the child's first step toward grammar, into territory beyond the reach of signing chimps and talking parrots. A viable account of this transition could thus both illuminate our theoretical understanding of human cognition and spur more practical efforts to simulate intelligent behavior.

But inquiries into the origins of such combinations have proven famously divisive. The crux of the debate concerns the role of meaning in grammatical theory and language learning — or, in more polarized terms, whether meaning plays any role at all. Conflicting stances on this issue have led to divergent assumptions in the literature about many core issues, including: what innate or prelinguistic biases children bring to the task, what kind of data constitutes the input to learning, how the target linguistic knowledge is represented, and to what degree language learning interacts with other linguistic and cognitive processes. The resulting array of disjoint research communities has yet to come to any consensus about the nature of the problem, let alone how best to address it.

This work explores and formalizes the view that grammar learning, like all language learning, is driven by *meaningful language use*. I take the following three claims as foundational:

- The basic linguistic unit is a pairing of form and meaning, *i.e.*, a **construction**.
- Meaning is **embodied**: it is grounded in action, perception, conceptualization, and other aspects of physical, mental and social experience.
- Language learning is **usage-based**: language structure emerges from language use.

In line with these claims, I propose a formulation of the grammar learning problem that gives a central role to meaning as it is communicated in context: the target of learning is a construction; the input to learning likewise includes not just utterance forms but also meaningful situation descriptions; and learning biases and strategies are designed to facilitate successful communication.

The focus of this investigation is on the emergence of structure: that is, how are multiple forms and meanings combined into larger linguistic units? These units run the gamut from fixed word sequences (*e.g.*, *fall down*) to partially abstract **item-based constructions** (Tomasello 1992) (*e.g.*, *N throw N*), through the more abstract patterns uncontroversially thought to constitute grammar (*e.g.*, argument structure constructions). Crucially, all of these can be seen as having underlying constituent and relational structure, giving rise to technical challenges not typically present in single-word learning: children must learn to associate *relations* in the forms they hear (such as word order) with *relations* in their accompanying meanings (such as event and participant structure). I refer to these associations as **relational constructions**.

This framing of the problem serves as the basis for a computational model of how relational constructions can be represented, used and learned. The model appeals to the intuition that children learn new constructions that improve their ability to make sense of the utterances they hear, and they do so based on experience. I propose a linguistic formalism suitable for capturing constituent structure and relational constraints. This formalism plays a central role in two processes, as shown in Figure 1.1: language understanding, which uses constructions to interpret utterances in context; and language learning, which seeks to improve comprehension by making judicious changes to the current grammar in response to input examples. Together these provide a computationally precise means of exploring the space of possible grammars, a quantitative basis for explaining the shifting dynamics of language acquisition, and a useful experimental platform for testing ideas from the developmental literature. When applied to input corpora of child-directed utterances paired with contextual descriptions, the model exhibits learning behavior compatible with crosslinguistic developmental patterns.

Even at this high level, it should be clear that we are venturing into complex and contentious territory. As with any cognitive modeling enterprise, the incomplete state of our current knowledge and the sheer number of structures and processes involved necessitate many simplifications and assumptions. Issues related to language acquisition boast an especially colorful history, having launched several disciplines' worth of clashing methods, goals and terminological distinctions. I thus begin by establishing some common ground in this chapter. Section 1.1 provides an informal sketch of the key ideas to be explored, illustrating the basic workings of the model using a simple example. Section 1.2 supplies some historical context to locate and motivate the proposed model with respect to the longstanding debate over the nature of grammar learning as well as more recent developments that have led to the approach pursued here. Finally, Section 1.3 summarizes the route to be taken as we delve into the details.

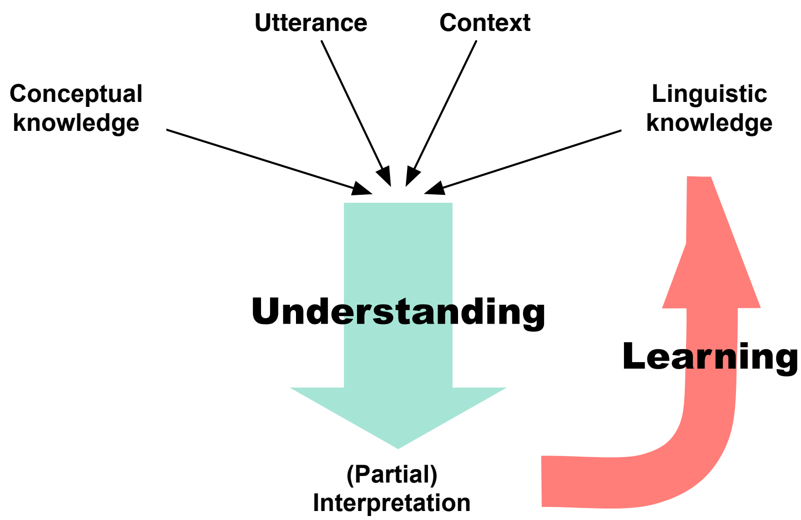

Figure 1.1. Learning by partial understanding: A schematic diagram of how the language understanding process draws on both conceptual and linguistic knowledge to produce an interpretation of a specific utterance-context pair. Partial interpretations can prompt learning processes that update linguistic knowledge.

1.1 Meaningful language learning: the model in brief

Our story begins with the child on the cusp of the two-word stage, in the latter half of the second year. By this point she has passed many milestones in the fitful climb toward full-fledged conversation. She has long since mastered how to point, reach and even cry her way to getting what she wants, and she has discovered that certain intonations and sounds (like her name, or a sharply rebuking "no!") merit special attention. Sometime around the nine-month mark, she began making sounds recognized by her eager parents as bonafide words; since then she has amassed a vocabulary that includes names for familiar people, objects and actions, as well as words associated with social routines ("hi", "bye-bye"), requests ("up", "more") and other complex events. Along with these overtly linguistic achievements, she has organized the flow of her experience into a stable

Age	Parent	Child
1;6.16	want to go see the snow?	look.
1;6.19	You lie down too, with the baby.	
	Put your head down.	
	Yes # that's a good girl.	
1;8.0	Nomi put the flowers in the other thing.	
	Don't throw them on the ground.	
	Put them right in here.	
	That's the way to do it.	*(unintelligible)*
1;11.9	They're throwing this in here.	
	throwing the thing.	throwing in.
		throwing in.
		throwing.
	throwing the frisbee.	
	...	
	What do you do with a frisbee?	frisbee.
	Do you throw the frisbee?	
	What do you do with a frisbee?	frisbee.
	Do you throw it?	throw it.
		I throw it.

Table 1.1. Example parent-child exchanges from 1;6.16 to 1;11.9 (Naomi corpus, Sachs 1983). (The age notation indicates the number of years, months and days.)

conceptual universe, acquired routine ways of engaging with different categories of objects and people and become a skillful reader of the intentions of those around her.

All of this expertise, linguistic and non-linguistic, has equipped her to make sense of the utterances and events around her, even when what she hears far outstrips in complexity what she can say. For concreteness, we can consider some typical parent-child interactions recorded during this period for the English-speaking child Naomi (Sachs 1983; MacWhinney 1991), shown in Table 1.1. These demonstrate the expected progression on the child's part from rare and tentative contributions (as in the first few segments) toward more active verbal participation resembling a conversation, with clear turn-taking and more coherent replies (as in the last segment). Despite the relative complexity of the parent's utterances throughout, the child often responds appropriately to the situation around her, in action if not in word. Even the earlier lopsided exchanges show signs of successful communication, based on either the adult's comments (*e.g.*, "Yes, that's a good girl", "That's the way to do it") or the child's fragmentary (but not merely imitative) contributions.

The current work assumes that the child's primary goal is to (re)act appropriately in this sense, and that all other developments, linguistic and non-linguistic, subserve this larger purpose. Early on, the child's apparent ease as an event participant depends less on her tenuous grasp of language than on her still-developing domain-general sensorimotor, cognitive and social skills. But over time, specific sound patterns may recur with specific entities and actions often enough that they become increasingly reliable cues directing her attention to those entities and actions. These correlations eventually become so entrenched that the sound patterns themselves directly activate the corresponding concepts, thus guiding and constraining the child's interpretations of utterances containing these patterns even in the absence of direct contextual support.

The process just described is consistent with many models of how the meanings of concrete terms like *frisbee* and *throw* are learned. These mappings — or **lexical constructions** — may be initially tied to particular items or contexts. But gradually the child becomes confident upon hearing "frisbee" that she should attend to a nearby frisbee, or retrieve her favorite frisbee from the next room, or imagine an appropriate frisbee-like referent. Likewise, the sound "throw" may initially refer to a specific configuration of actors and projectiles, but over time it becomes a reliable cue to attend to or expect a particular motor pattern performed by any actor on any object. These lexical mappings provide ever more useful constraints on the child's understanding of the situation.

My main interest in this work is what happens as the child begins to encounter multiple such associations within individual utterances. Having acquired lexical constructions for *frisbee* and *throw*, for example, a child could infer that the utterance "Do you throw the frisbee?" from the excerpt at 1;11.9 (*i.e.*, at the age of 1 year, 11 months and 9 days) may have something to do with both a throwing action and a frisbee; pragmatic constraints might further lead her to conclude correctly that the frisbee is the thrown object in this potential scene. Note also that the words appeared together in a sentence, in a particular order.

I hypothesize that the same general mechanisms supporting lexical learning support the acquisition of more complex constructions. That is, at all stages of learning, the child draws on her entire conceptual, pragmatic and linguistic arsenal to infer the intentions of the people around her and cope with utterances beyond her own productive capacities. Aspects of the speech signal that reliably correlate with aspects of the situation become prime candidates for new linguistic mappings. The main difference as the child moves toward multiword constructions is that these mappings rely on previous mappings, and more specifically involve *relations* linking existing mappings. In our example scenario, the form relation "*throw* comes before *frisbee*" may become associated with the meaning relation "throw action with frisbee as object being thrown". These relational correspondences between form and meaning are similar in kind to simpler lexical constructions but structurally more complex. Over time and repeated use, such associations become increasingly entrenched; they may be combined into larger or more general mappings.

The relational nature of these complex constructions is the source of the major challenges addressed by the model, namely: how are relational constructions represented, used and learned? Although we have touched on each of these issues, the full story is much more complex. The initial formation and later reorganization of relational constructions is driven directly by (partial) language comprehension; it also relies on a quantitative metric to choose an optimal set of constructions that allows the child to generalize well to new situations while recognizing statistically common patterns. All of these processes and the structures upon which they operate must, moreover, be described in computationally precise detail. But the informal sketch provided here of the key issues at hand should serve as sufficient initial orientation to the model.

1.2 Goals in context

We turn now to the broader theoretical context against which to define the goals of the current work. A satisfactory account of the human capacity for language can be divided into three distinct components (Chomsky 1986):

1. language structure: the defining structural properties of natural languages
2. language acquisition: how children acquire their native language
3. language use: how linguistic and nonlinguistic knowledge interact in comprehension and production

This division raises a theoretical chicken-and-egg problem. On the one hand, the structure of (1) can be seen as "logically prior" (Chomsky 1957) to the processes of (2) and (3): a theory of acquisition depends on what is being acquired, and a theory of use depends on what is being used. On the other hand, language is as language does: we can glean clues about the nature of linguistic structure only indirectly, by observing its associated behaviors — comprehension and production, in adults and over the course of development. From this perspective, the theory of language structure is constrained to be functionally adequate to support these observed behaviors: it must be both learnable and usable.

This thesis offers a computational account of how meaningful language use drives the transition from single words to complex combinations. Both the problem definition and its proposed solution are motivated by two interdependent goals. The first of these is scientific: the account should be maximally consistent with all available evidence relevant to the theoretical components above — language structure, acquisition and use — and offer concrete realizations of each of them. Equally important, however, is a methodological goal of showing how the origins of grammar, like other challenges in cognitive science, may be fruitfully addressed by integrating constraints from multiple perspectives and disciplines. I focus on combining ideas and methods from developmental psychology, linguistics and computer science to triangulate on a model that is both cognitively plausible and technically precise.

The resulting model differs markedly from approaches elsewhere in the literature, both in the breadth of evidence it is intended to encompass and its multidisciplinary pedigree. In this section I first outline the kinds of linguistic evidence that will be considered relevant and summarize the main relevant assumptions (Section 1.2.1). I then situate the model within its background research context of the Neural Theory of Language project (Section 1.2.2), which seeks to illuminate how cognitive and linguistic behaviors are realized using mechanisms of neural computation. Finally, I return to articulate the specific goals and contributions of the current model with respect to the main disciplines involved (Section 1.2.3).

1.2.1 Grammaticality and its discontents

What is the nature of linguistic knowledge? Chomsky pioneered the use of *grammaticality* judgments as a prime source of evidence in this matter. Most speakers of English, for example, deem (1–1a) grammatical, (1–1b) grammatical (if nonsensical) and (1–1c) ungrammatical:[1]

(1–1) a. Rabid white dogs bite furiously.
b. Colorless green ideas sleep furiously. (Chomsky 1957)
c. *Sleep ideas colorless furiously green.

Such judgments have played an indisputable role in endowing linguistic inquiry, previously a largely descriptive endeavor, with something akin to the rigor of an experimental science. As Chomsky observed, the acceptability of examples like (1–1b) suggests that grammaticality requires neither previous experience with a sentence nor a meaningful interpretation. Evidence of this kind has allowed linguists to identify a wide range of structural and distributional regularities within and across languages, such as the syntactic similarities between (1–1a) and (1–1b).

But binary grammaticality judgments tell only part of the story. In contrast to the cases above, it appears that the acceptability of some sentences may be contingent on their interpretation, as demonstrated by the examples below:

(1–2) a. Pat sneezed the napkin off the table. (Goldberg 1995)
b. ? Pat laughed the napkin off the table.
c. ? Pat slept the napkin off the table.
d. ? Pat sneezed the boulder off the table.

[1] The traditional asterisk * marks grammatically unacceptable sentences, and ? marks sentences whose acceptability is questionable or variable.

Whether these sentences are grammatically foul or fair is subject to question. The differences in acceptability are not due to syntactic notions like part of speech, since all of these involve transitive uses of canonically intransitive verbs, nor to the lexical semantic properties of the verb alone, since (1–2a) and (1–2d) differ only in their object. Rather, the fact that (1–2a) yields a more coherent interpretation than the others hinges on both the presence of a directed force, albeit a nonstandard one, in the act of sneezing (as opposed to laughing or sleeping); and the relative susceptibility of napkins (as opposed to boulders) to being propelled by such a rhinal gust.

As discussed further in Chapter 2, Goldberg (1995) cites such examples as evidence that the syntactic structure itself imposes semantic requirements that must be collectively satisfied by the relevant arguments — requiring, in this case, that Pat's action cause the relevant entity to move off the table, or more generally, that there is a causal relationship between the action and the motion. Goldberg identifies the particular pairing of syntactic and semantic constraints observed here as the CAUSED MOTION construction: sentences with this syntactic structure are acceptable to the degree to which such a caused-motion reading is available, such that even the anomalous cases in (1–2) may be rendered less so given some ingenuity in devising contextual conditions. Goldberg further analyzes the required event as a conceptual category with prototype structure like that observed for other cognitive and linguistic categories (Lakoff 1987), where the central case of caused motion involves the direct application of physical force. Thus, while canonically transitive verbs appear most frequently and naturally in this construction, other verb-argument combinations that can be construed as caused motion are also permitted. The examples in (1–3) below draw on other kinds of causation, such as the curative power of rest in (1–3a), the social-emotional dynamics of derision in (1–3b), or even an abstract or metaphorical economic force like that in (1–3c):

(1–3) a. Pat slept the cold out of her system.
b. Simon laughed the singer off the stage.
c. The weak dollar pushed the country into a recession.

Note that the acceptability of these examples without any additional contextual support (unlike the marked cases in (1–2)) may derive in part from idiomatic uses of the verbs involved, as in the expressions *sleep your troubles away* and *laugh him out of the room*, or from a lexicalized abstract sense of *push*. Even these uses, however, appear to be constrained by the same essential semantic requirement of a caused-motion scene, in which the specified action can plausibly be construed as effecting the specified motion.

What do such seemingly exotic phenomena tell us about linguistic structure? And how do these complexities relate to the problem faced by the child? Evidence from children's use of language suggests that the examples above have counterparts at the earliest stages of language learning. The overgeneralizations cited in Bowerman (1988) below, for example, though not syntactically well-formed by adult English standards, were nevertheless semantically transparent in context:

(1–4) a. ? She came it over there. *(≈ She made it go over there.)*
b. ? Will you climb me up there? *(≈ Will you help me climb up there?)*
c. ? Don't giggle me. *(≈ Don't make me giggle.)*

Both (1–4a) and (1–4b) can be viewed as instances of the CAUSED MOTION construction discussed above, and (1–4c) as an instance of a different causative construction associating transitive syntax with a caused action, as in *Don't move me* (*i.e.*, Don't make me move). Like the earlier examples, such creative language use underscores the close connection between syntactic patterns and the meanings with which they are associated. The syntactic pattern may license nonstandard or extra verbal arguments that fulfill its semantic conditions.

Some prominent theories of language discount the phenomena like those discussed here as due to context-dependent vagaries of *performance*, or as poetic, peripheral or idiomatic language use with little bearing on core principles of grammatical *competence*. These core principles are restricted to relatively abstract, structural properties of language for some thinkers, notably Chomsky, who suggested that "semantic notions are really quite irrelevant to the problem of describing formal structure" (Chomsky 1955, 141, cited in Huck & Goldsmith (1995)). On this view, grammar is assumed to be *autonomous* from meaning and use, with evidence tilted heavily toward binary grammaticality judgments like those of (1–1). This focus on structural and symbolic aspects of language has had far-reaching consequences, especially for the study of language acquisition. In particular, as described further in Chapter 2, the language learning problem as defined within this framework turns out to be provably impossible (Gold 1967), leaving developmental theorists in this paradigm no recourse but to assign a correspondingly greater role to innate language-specific knowledge.

The position taken here is less restrictive. I regard the nature of linguistic knowledge as a sufficiently open question to warrant the inclusion of all kinds of evidence. In particular, the phenomena illustrated in the examples above hint at a sprawling grammatical grey zone between the extremes of unequivocal acceptance or rejection. Grammaticality judgments appear to depend on a complex interaction between the sentence as a whole and its constituent parts, which in turn cannot be insulated from semantic and pragmatic considerations, encyclopedic world knowledge and distributional factors affecting degree of idiomaticity — in short, from all aspects of meaning and use. These ideas are consistent with a growing body of research suggesting that the study of formal linguistic structure cannot be so easily segregated from facets of meaning and function. These include typological studies of the range of crosslinguistic variation; historical evidence of the nature of language change; psycholinguistic experiments measuring behavioral and neurobiological correlates of language use; and, most relevantly for the current enterprise, crosslinguistic developmental research on how children learn and use language. All of these support the idea that the phenomena discussed above, far from being quirks of a particular expression, may instead reflect inherent characteristics of linguistic units of all kinds and across languages.

The foregoing observations bear directly on the choice of a suitable representation for the complex relational constructions that are the topic of this thesis. I focus here on evidence most relevant to how words combine to form larger meaningful units, and how such combinations might be learned, as summarized by the foundational assumptions stated at the outset of this chapter:

- **Constructions**: The basic structural unit of language is a mapping between form and meaning, or **construction** (Goldberg 1995; Kay & Fillmore 1999; Langacker 1987; Croft 2001; Goldberg & Jackendoff 2004). No strict division between grammar and the lexicon is assumed: both form and meaning constraints may be associated with units of all sizes and at all levels of abstraction. The meaning of a complex multi-unit construction need not be a strictly compositional function of the meaning of its parts.

- **Embodiment**: All concepts are grounded in human conceptualization and embodied experience. Linguistic structures thus reflect principles of cognitive organization: they are grounded in action, perception, emotion and other aspects of experience, and they are subject to categorization effects (Lakoff 1987; Langacker 1987; Talmy 2000).

- **Usage**: Language structure emerges from language use; conventions of linguistic knowledge emerge from the totality of entrenched usage events (Langacker 1987; Bybee 2001). Language acquisition is likewise linked directly to specific usage events and driven by the statistical properties of language use (Slobin 1985; Tomasello 2003; Clark 2003).

Each of these strands is independently well motivated by the evidence, reviewed in Chapter 2. Together they constitute a coherent and multifaceted challenge to traditional approaches to language acquisition, with radically different conceptions of both the nature of the human capacity for language and its developmental course. This overarching framework holds great promise for illuminating crosslinguistic developmental phenomena while accommodating convergent evidence from multiple disciplines.

The current model does not attempt to deliver on that promise in its entirety; it aims rather to illuminate, from several perspectives, a well-defined subpart of the larger problem. In particular, I will focus on the representational needs of the earliest relational constructions, and how they differ from simple lexical constructions. Nevertheless, many of the background assumptions and technical aspects of the model are motivated and constrained by these long-term goals, as articulated within the broader research context of the Neural Theory of Language project. I next review the key ideas and forerunners of the current model as defined within that framework.

1.2.2 Embodied models of language acquisition and use

The work described in this thesis evolved in the context of the Neural Theory of Language (NTL) project, an interdisciplinary effort at the International Computer Science Institute and the University of California, Berkeley, to build models of language acquisition and use that are consistent with linguistic, biological, psychological and computational constraints. Feldman (2006) provides a comprehensive overview of the project's efforts toward answering the question: How does the brain compute the mind?[2] More specifically, how might the linguistic and psychological phenomena described in the previous section be instantiated in the human nervous system? How can models of cognition and language satisfy the computational constraints of the human brain?

The NTL project takes a layered approach to bridging the considerable gap between biology and behavior. Despite advances in imaging and other experimental techniques, it is not yet feasible, nor very informative, to build theories and models directly linking neural structures with language and cognition. Instead, we rely on computational modeling as a means of articulating and testing hypotheses. We have found it useful to distinguish several intermediate levels of abstraction, as shown in Table 1.2. The overall goal is to build models that are informed and constrained by theoretical proposals and empirical findings from each level. This framework encompasses the application of computational methods to model both cognitive and neurobiological phenomena, with the additional requirement that all computational mechanisms be reducible to mechanisms at more biologically motivated levels of description. It thus differs somewhat from many approaches to cognitive modeling, in which computational or algorithmic levels deliberately abstract away the underlying biological details (*e.g.*, Marr's (1982) computational level, or the ACT-R model (Anderson 1993)); it is also focuses on higher-level phenomena than computational neurobiological models. The key bridging level of description is that of *structured connectionist modeling* (Feldman & Ballard 1982), which shares the principles of neural computation that characterize connectionist modeling (Rumelhart *et al.* 1986; Elman *et al.* 1996) but emphasizes the inclusion of larger-scale structure like that found in neural architecture. This layered methodology makes it theoretically possible to re-

[2]See also `http://www.icsi.berkeley.edu/NTL`.

Cognition and language	cognitive mechanisms and linguistic phenomena
Computation	formalisms, data structures, algorithms
Structured connectionism	distributed networks of units
Computational neurobiology	models of neuronal structures and processes
Biology	biological/neurophysiological structures and processes

Table 1.2. Neural Theory of Language levels of description

spect constraints of all kinds while targeting a particular set of phenomena either within a level or across one or more levels.

Most of the early work in the group focused on the upper three levels and the connections among them. Regier's (1996) model of the crosslinguistic acquisition of spatial relations terms remains the paradigm illustration of how a structured connectionist model incorporating biologically plausible constraints — in particular, features of the human visual system — can be used to model high-level linguistic phenomena. The model addresses the original motivating problem for the NTL effort (then known as the L_0 project, where 0 refers to the approximate percentage of natural language covered): how can spatial relations between basic geometric shapes (such as triangles and squares) like English *above*, *below*, *in* and *on* (Feldman *et al.* 1996) be learned? More generally, how can universal (but non-linguistic) constraints give rise to a typologically diverse range of systems for labeling spatial relations?

Subsequent models have expanded both the range of cognitive and linguistic phenomena addressed and the repertoire of computational modeling tools available. Those most relevant to the current work fall into two categories: (1) simulation-based models of language understanding and inference that draw on dynamic, embodied representations of actions and events; and (2) statistical frameworks that support cognitively plausible models of language acquisition. These will be described in more technical detail in Chapter 2; here I describe each line of work informally and highlight its relation to the current model.

The first line of work explores the role of **simulation** (or imagination) as a core component of human intelligence and a cornerstone of NTL models of language use, as expressed below:

Simulation Hypothesis

Language exploits many of the same structures used for action, perception, imagination, memory and other neurally grounded processes. Linguistic structures set parameters for simulations that draw on such embodied structures.

The Simulation Hypothesis is consistent with abundant evidence of embodiment in language and cognition (see Chapter 2). It proposes an explicit link between linguistic structures — *i.e.*, constructions — and other independently motivated sensorimotor and cognitive processes. This view of the function of language offers significant representational and computational advantages. Crucially, constructions are relieved of having to anticipate every possible interaction or condition that may affect interpretation, and can rely instead on simulation to shoulder most of that semantic and pragmatic burden. On this view, an utterance is acceptable if it is licensed by a set of constructions that can parameterize a valid simulation in the current context, and its meaning consists of the set of inferences produced by this simulation.

The resulting active, embodied and context-dependent approach to linguistic meaning offers elegant solutions to longstanding problems, such as those highlighted below:

(1–5) a. ? Pat sneezed the {napkin, boulder} off the table.
b. The {hobo, smoke} drifted out of the house. *(portal = door, window)*
c. Mickey ran into {the house, the wall}. *(path vs. contact sense)*
d. Mary {swam, sneezed} for five minutes. *(single-event vs. iterative reading)*
e. The economy moved along at the pace of a Clinton jog.

As noted in Section 1.2.1, the acceptability of (1–5a) may depend on the nature and weight of the putative projectile. Event details — such as the most likely portal of exit in (1–5b) — likewise depend on material characteristics of the participants, as does the intended sense of the phrase *ran into* in (1–5c). Many problems in aspectual inference and coercion, like those in (1–5d), have also been analyzed as emerging from the dynamic combination of generalized event structures and the sensorimotor constraints of particular actions and objects (Narayanan 1997a; Chang *et al.* 1998). In all of these cases, contextually available information about the trajector (for example, that Harry is a mouse or a ghost) or a special background context (for example, watching a film in slow motion) may render otherwise dispreferred readings acceptable. Finally, as described by Narayanan (1997a), metaphorical expressions like (1–5e) can be understood as cueing simulations in an embodied source domain whose inferences are mapped onto a target domain. Such inferences are unlikely to be lexicalized or otherwise linguistically encoded. Rather, they arise from the dynamic composition of linguistic and ontological constraints in context.

The other relevant line of NTL research has focused on developing models of language learning that are both statistically grounded and representationally adequate for a variety of domains. The Regier spatial relations model relied on backpropagation learning in a structured connectionist model; later models have applied probabilistic and information-theoretic techniques to model different aspects of acquisition under more realistic learning conditions. Stolcke (1994) established a general framework based on *Bayesian model merging* (Stolcke 1994; Stolcke & Omohundro 1994) for learning probabilistic formal grammars from relatively few examples; these included probabilistic attribute grammars suitable for generating simple descriptions of the L_0 spatial relations domain (*e.g.*, *A circle is above a square*). Bailey (1997) applied the same model merging approach to the more cognitively motivated domain of hand action verbs, showing how the lexical semantics of *push* and *shove* in English, as well as comparable data from other typologically distinct languages, can arise from appropriately parameterized, embodied representations. Both the algorithm and these most relevant applications are described in detail in Chapter 5.

This thesis brings together these separate efforts in the language understanding and language learning spheres. The linchpin of this union is a grammatical formalism that redresses a representational shortcoming for both domains. The understanding models have been concerned with rich embodied representations for simulative inference, and have not explained how linguistic structures fit together to parameterize these simulations. The learning models have focused on relatively restricted domains of both form (either single words or simple syntactic patterns) and meaning (semantic features, provided in or extracted from the input), as required to make learning tractable. In short, neither strand of work has embraced the cross-domain, relational structure found in construction-based approaches to grammar.

The model presented here remedies these limitations with an embodied, usage-based construction grammar formalism consistent with the foundational assumptions set out in Section 1.2.1. This formalism satisfies the representational needs of both understanding and learning: it serves as the link between complex linguistic structures and the embodied schemas used to parameterize simulations; and it provides a linguistically more adequate target of learning than previous learning models. The learning methods used in earlier models are extended to handle the structural demands of the formalism and to exploit the processes of language use, which in turn are able to accommodate increasingly more complex expressions.

1.2.3 Contributions

The model to be described has three interdependent components, each of which integrates psychological, linguistic and computational constraints:

- an embodied construction grammar formalism;
- a simulation-based model of language understanding; and
- a usage-based construction learning algorithm.

My priority is to model the emergence of linguistic structure as faithfully and precisely as possible. As in many cognitive modeling tasks, concerns relevant to individual disciplines will generally be subordinated to this larger goal. To ease exposition, I will opt for examples that highlight the core computational problems addressed; these examples may appear cartoonishly oversimplified to linguists used to more syntactically complex and semantically nuanced territory, and they may capture only a subset of the factors psychologists have identified as impinging on the child's learning problem. Computer scientists, by contrast, may question the meaning-oriented approach taken here, or deem it prematurely restrictive to adopt human memory and processing limitations.

If such concerns can be suspended, however, the multidisciplinary perspective offers many benefits not otherwise available to the individual research communities involved. Above all, the model presented here helps to validate a constructional, embodied, usage-based approach to language. These ideas depart significantly from some widely held assumptions about the nature of language (discussed further in Chapter 2) and are often dismissed as insufficiently well-specified or constrained to be feasible. In demonstrating that these alternative assumptions can be formalized, and further that they support working models of comprehension and acquisition, the model serves as an existence proof that may quell some of these criticisms. More importantly, computationally precise descriptions of the structures and processes proposed should facilitate concrete discussion among formally minded researchers from all disciplines.

The model may also have a broader unifying function across subdisciplines in the study of language. As befits the current focus on early child language, many of the illustrative examples used are based on relatively simple phenomena that exhibit the kind of relational, structural complexity described earlier. The formalism to be presented is, however, designed to scale to more complex cases and accommodate representational needs that crosscut traditional divisions. It thus has the potential to provide a unified framework for studying linguistic patterns of all kinds — synchronic and diachronic, in child and adult language, idiomatic and grammaticized, crosslinguistic and language-internal — and collected in all ways: in all modalities, in spontaneous and elicited conditions, measured with the tools of psycholinguistics or neurobiology. The study of these phenomena has been split by historical tradition and accident across several academic disciplines; I hope that a more catholic representation of their common object of study will reveal recurrent themes and permit more integrated analyses.

From a computational linguistic perspective, the model establishes a new problem space for theoretical and practical investigation, based on previously unformalized assumptions about language. This space poses some representational challenges not typically addressed in tasks like speech recognition, part-of-speech tagging and syntactic parsing; indeed, the remarkable progress in these areas over the last decade has been driven by the rise of statistical methods and large-scale speech and text corpora, and researchers in these areas have largely excluded explicit meaning representations and are wary at best of linguistic theory.[3] But problems more akin to human behaviors — such as language understanding and language acquisition — have proven more resistant to such techniques. Such problems are essentially communicative in nature, and thus may require the formalization of notions like meaning, context and intent. It seems reasonable that our computational methods should draw inspiration from their most salient counterparts in human cognition.

Finally and more technically, the learning problem described here instantiates a more general class of problems worthy of study independent of its application to language. The representational structure posited for language involves two domains (form and meaning), with internal relational structure in each domain as well as structured correlations across the two domains. Such coupled relational representations may be an appropriate formalization for a wide array of problems, cognitive and otherwise. In this vein, the model explores how principles of information theory and probability theory can be applied to learn such relational representations from sparse data.

1.3 Road maps

With every utterance, a speaker forces all the sprawling, nuanced and richly interconnected aspects of some communicative intention into a finite sequence of forms. This thesis likewise linearizes an inherently non-linear set of ideas into the chapters that follow. I here summarize the main narrative flow and provide some suggestions for readers who wish to choose their own adventures.

Chapter 2 is intended to level the playing field across a multidisciplinary audience; it includes a brief overview of linguistic, developmental and computational theories of language, as well as a primer on the standard course of acquisition. Motivating evidence most pertinent to the model is organized around the foundational assumptions stated at the outset of this chapter. Readers familiar with the relevant background may safely skip through this chapter; specific assumptions made by the model will be reviewed in subsequent chapters.

Chapter 3 presents the Embodied Construction Grammar formalism that serves as the hypothesis space of the learning model. While the discussion is geared toward aspects of the formalism most relevant for learning, it is also essential that the formalism chosen has the potential to fit into the larger theoretical framework, as well as express more complex constructions. I thus delve into considerable linguistic detail in describing the schema and construction formalisms, as well as the simulation-based model of language understanding it supports. For non-linguists, the overview in Section 3.1 may suffice as an introduction to the goals, representation and function of the formalism.

[3]Note the implied causal link in Jelinek's (1988) remark "Every time I fire a linguist our system performance improves."

Readers wishing to cut to the computational chase may proceed directly to Chapter 4, which along with Chapters 5–7 constitutes the technical core of the thesis. Chapter 4 recasts the ideas of Chapter 3 in more technical terms: concepts relevant to both understanding and learning are formally defined; algorithms for interpreting utterances and resolving references in context are described; and various evaluation criteria for assessing these interpretations are specified.

Chapter 5 serves as a key bridging chapter in several respects. Having proposed provisional theories of language structure and use in the preceding two chapters, I shift the focus more directly to learning, synthesizing the foregoing developmental and linguistic constraints into a formal statement of the learning problem and sketching out the general solution pursued in this work. This chapter expands the high-level story told in Section 1.1 to a conceptually more precise but still relatively non-technical level of description.

The next several chapters supply the technical and algorithmic specifications of the learning model, instantiating the general approach set out in Chapter 5 and applying it to experimental data. Chapter 6 defines several usage-based learning operations for proposing new constructions and adapting existing constructions based on new input examples. These operations divide broadly into the two categories of context-driven relational mapping and constructional reorganization. Chapter 7 specifies the quantitative evaluation metric for choosing which of these moves through the space of possible grammars to make. The proposed optimization-based criteria are designed to minimize the cost of encoding both the current grammar (*i.e.*, the set of constructions) as well as any observed data in terms of the current grammar. Chapter 8 describes results of applying the ideas in the preceding chapters to a corpus of input data based on transcripts of child-directed language. Experimental results demonstrate that the model can acquire simple relational constructions, including both concrete lexical constructions and item-specific constructions exhibiting varying degrees of generalization.

Finally, Chapter 9 looks onward and outward, drawing connections to related research and suggesting directions for continuing the work begun here. I conclude by considering the implications of the model and the ideas it advances.