Constructing grammar: A computational model of the emergence of early constructions

Nancy Chih-Lin Chang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2009-24
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-24.html

February 5, 2009

Constructing grammar:
A computational model of the emergence of early constructions

by

Nancy Chih-lin Chang

A.B. (Harvard University) 1994
M.Phil. (Cambridge University) 1995

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy
in

Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Jerome A. Feldman, Chair
Professor Nelson Morgan
Professor Daniel I. Slobin

Fall 2008

The dissertation of Nancy Chih-lin Chang is approved.

Chair		Date

Date

Date

University of California, Berkeley

Fall 2008

Constructing grammar:

A computational model of the emergence of early constructions

Abstract

Constructing grammar:
A computational model of the emergence of early constructions

by

Nancy Chih-lin Chang
Doctor of Philosophy in Computer Science

University of California, Berkeley
Professor Jerome A. Feldman, Chair

In this thesis I explore and formalize the view that grammar learning is driven by meaningful language use in context. On this view, the goal of a first language learner is to become a better language user — in particular, by acquiring linguistic *constructions* (structured mappings between form and meaning) that facilitate successful communication. I present a computational model in which all aspects of the language learning problem are reformulated in line with these assumptions. The representational basis of the model is a construction grammar formalism that captures constituent structure and relational constraints, both within and across the domains of form and meaning. This formalism plays a central role in two processes: language understanding, which uses constuctions to interpret utterances in context; and language learning, which seeks to improve comprehension by making judicious changes to the current set of constructions.

The resulting integrated model of language structure, use and acquisition provides a cognitively motivated and computationally precise account of how children acquire their earliest multiword constructions. I define a set of operations for proposing new constructions, either to capture contextually available mappings not predicted by the current grammar, or to reorganize existing constructions. Candidate constructions are evaluated using a minimum description length criterion that balances a structural bias toward simpler grammars against a data-driven bias toward more specific grammars. When trained with a corpus of child-directed utterances annotated with situation descriptions, the model gradually acquires the concrete word combinations and item-based constructions that constitute the first steps toward adult language.

Professor Jerome A. Feldman
Dissertation Committee Chair

for my parents, I-Cheng and Phoebe Chang

A word after a word
after a word is power.
— *Margaret Atwood, from "Spelling"*

We are all talkers
It is true, but underneath the talk lies
The moving and not wanting to be moved, the loose
Meaning, untidy and simple like a threshing floor.
— *John Ashbery, from "Soonest Mended"*

Contents

List of Figures

List of Tables

Acknowledgments

I have many blessings to count.

Most of these are centered around the research haven that is the International Computer Science Institute, radiating out to encompass the Berkeley campus at large. I owe the comforts of my niche here, both intellectual and practical, to Jerry Feldman. His vision and guidance as a leader of the Neural Theory of Language project have been fundamental to the interdisciplinary synthesis undertaken in this work, and I am grateful to him for creating a research environment that encourages such free thinking about such heady topics. And if patience is a virtue, he has achieved true beatitude as my adviser, doing whatever he could to help with matters large and small. I hope it will have been worth the long wait.

Several others have also been indispensable in shepherding this work through its various stages. I am fortunate to have been enlightened early on by Dan Slobin's careful scholarship and unfaltering wisdom in the developmental domain. He and Nelson Morgan have shown true grit as members of my dissertation committee, and I thank them both for their willingness to engage across disciplinary divides. I am also grateful to Marti Hearst and Stuart Russell for providing useful feedback as members of my qualifying exams committee. Unofficially but just as crucially, George Lakoff's fearless thinking and infectious enthusiasm have influenced me deeply, and Srini Narayanan has never hesitated to offer his counsel and encouragement in his roles as Font of All Knowledge, shadow adviser and friend. Many thanks also to Lokendra Shastri and Eve Sweetser for many enlightening exchanges over the years. Finally, I have been blessed by my proximity to the linguistic grace of Chuck Fillmore and Paul Kay. I hope I have gleaned something from my time among these sages.

Perhaps less saintly but no less appreciated have been my fellow pilgrims slouching toward the realization of our (im)modest NTL ambitions; the hilarity and insight they provided during late nights (and early mornings!) in the office made for an entertaining journey indeed. I could not have asked for a better officemate, collaborator and friend than Eva Mok: reflexively helpful, (almost) inhumanly competent, and somehow always equipped with whatever I needed, from research papers and debugging advice to office supplies and caffeine. I cannot imagine how I could have survived the final push without her (it turns out that misery *does* love company). I also hit the jackpot with my other ECG collaborators extraordinaire: my personal psycho-linguist Ben Bergen (who really does put the "fun" back in phonology, *inter alia*), with whom I first induced multiple hallway door closures; and Johno Bryant, the vampiric co-pilot of *la coche de la noche*, whose eponymous daily hour of comic stylings I shall sorely miss.

ICSI has brought a parade of sharp minds and kindred spirits from around the country and world to my door. I am indebted to Keith Sanders, Robert Porzel and Tiago Maia, for valuable conversations and collaborations that helped shape my thinking about language and learning; the inimitable Josef Ruppenhofer (my personal *Gastarbeiter*) for cheap (outsourced) labor; polyglots Patrick Chew, Eva Mok and Vanessa Micelli for data-wrangling assistance and consultation; the illustrious Michael Israel for pragmatic(s) advice; Ellen Dodge, for many fruitful discussions; and Joe Makin, Steve Sinha and the rest of the Friday night crowd for our joint philosophical investigations. Thanks also to Collin Baker, Hans Boas, Michael Ellsworth and my other FrameNet friends across the hall; my pals from the really old days, David Andre (fellow seeker of Big-I interesting), Dan Gildea (my personal computational linguist) and Carter Wendelken; and many other colleagues and collaborators who have contributed directly and indirectly to my research life, including: Alberto Amengual, Klinton Bicknell, Ingrid Fischer, Andrea Frome, Olya Gurevich, Colin Holbrooke, Robert Hunter, Teenie Matlock, Shweta Narayan, Mark Paskin and Nathaniel Smith. A few guardian angels who have cheered me throughout deserve special mention: Nikki Mirghafori, who has been much more than a departmental big sister to me; the ever-supportive Miriam Petruck; and Madelaine Plauché, who kept me supplied with articles, fresh herbs and hugs.

I am also fortunate to have followed in the hallowed footsteps of ICSI/Berkeley luminaries past and present. Thanks to David Bailey, Ben Gomes, Claudia Brugman, Adele Goldberg, Dan Jurafsky, Srini Narayanan, Terry Regier and Andreas Stolcke for dispensing encouragement and wisdom from near and far. I am especially indebted to Laura Michaelis for her many insights, linguistic and otherwise; and Susanne Gahl, for top-notch dissertation doula services and yak-defrosting advice.

Outside of ICSI, I am blessed with a collectively and individually polymathic, mirthful coterie of companions who have kept me sane (relatively) and entertained (absolutely). Pride of place and boundless appreciation go to Jason Horowitz, who took it upon himself to shove me across the finish line. He has helped me, heroically, in matters motivational, mathematical, technical, implementational, bureaucratical, logistical, even hysterical; I positively could not have done this without him. I am also unsure what I did in a former life to deserve the further phalanx of first-rate friends who have fed body, soul, mind and spirit: my longtime bedrocks Ambika Anand Prokop and Paul Prokop, who have been there for me (and with me) across oceans and continents, and whose List I am glad to hereby vacate; Jill and Marc Alvidrez, who have opened heart and home to provide sanctuary (and sustenance!) in many times of need; Vin de Silva, whose unflagging belief in me (and LaTeX-pertise) exceeded my own; the incomparable Huma Dar, my fellow traveler in every sense; and, most non-presumptuously, Ted Blanchard, who has supplied much-needed succor and inspiration (editorial, comical, metaphorical and more) during my lengthy captivity.

I am also grateful to the wider circle of friends, colleagues, housemates, travel companions and well-wishers who are responsible, variously, for periodic infusions of encouragement, gentle/stubborn prodding and diversions of all kinds. Thanks in particular to Harriet Chang (who tolerated my schedule longer than anyone else), David Baker (an eagle-eyed editor's editor), Richard Iles Tucker (my typographical guru), Mira Bernstein, Josh Tenenbaum, Sanjoy Dasgupta, Steve Ketchpel, Neeta Mhatre and Trevor Stockinger. To all of them and many others far too numerous to mention: here's hoping for more shared gastronomic, cultural and conversational delights, and further adventures in Moroccan deserts, Chinese "highways" and crumbling *ruinas* everywhere.

Finally, I would like to thank my parents, I-Cheng and Phoebe, and my sisters, Carrie and Sand, for all the love and support they offered, as well as all the lapses in filial and sororal attention they endured, with extra appreciation to Sand for sharing with me the ultimate creature comfort, my canine nephew Theo. I would also like to remember here my grandmother, Tze-Hwa Teng, through whom I first came to appreciate that language learning isn't as easy as one might think.

Any errors and omissions remaining in this work despite the litany of blessings above are, of course, my responsibility alone.

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.

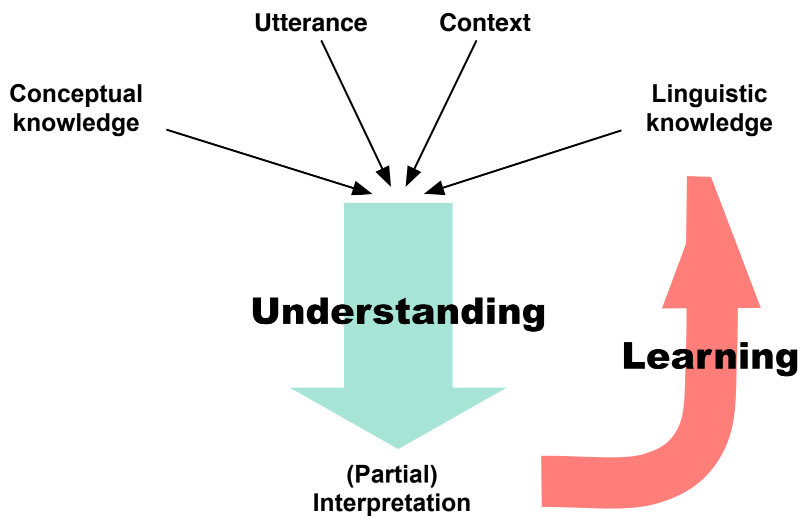

Figure 1.1. Learning by partial understanding: A schematic diagram of how the language understanding process draws on both conceptual and linguistic knowledge to produce an interpretation of a specific utterance-context pair. Partial interpretations can prompt learning processes that update linguistic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

Chapter 2

Orientations

Meanings cannot be defined in terms of our science and cannot enter into our definitions.
— Leonard Bloomfield, 1931

In language, forms cannot be separated from their meanings. It would be uninteresting and perhaps not very profitable to study the mere sound of a language without any consideration of meaning.
— Leonard Bloomfield, 1943

Of the above possible fields the learning of languages would be the most impressive, since it is the most human of these activities. This field seems however to depend rather too much on sense organs and locomotion to be feasible.
— Alan Turing, 1948

This chapter provides a highly selective primer on research relevant to the current model. Although these contributions are drawn broadly from the psychological, linguistic and computational realms, the boundaries separating these fields are porous and the interconnections deep. Indeed, the peculiar history of the scientific study of language and language learning, and especially the role of meaning in these endeavors, has given rise to parallel developments and rifts in each of these fields. As captured by the insights from Bloomfield and Turing above, these reflect both the conviction that language as an object of study must involve meaning and the conceit that the tools of "our science" seem inadequate to the task. The main divisions in each of the relevant disciplines might be seen as embodying two responses to the resulting quandary: either limit the object of study to a domain for which our scientific tools are adequate; or sharpen our tools to address the original object of study.

Despite the pessimistic tone struck by the progenitors of modern linguistics and computer science, I will suggest that prospects for including meaning in our theories have improved in the interim. Section 2.1 presents a brief review of language acquisition research, tracing it from the first formal statement of the problem to its current divided state. Section 2.2 then highlights diverse research supporting constructional, embodied and usage-based views of language. Computational concepts and tools that play a key role in the model are summarized in Section 2.3.

2.1 Preliminaries

2.1.1 Gold's legacy

Language acquisition has served as a natural route into studies of the mind for thinkers from Aristotle and Augustine onward. Only relatively recently, however, has the problem lent itself to mathematical formalization. The key catalyzing influence was the Chomskyan program in its various incarnations (1957,1963), which uncovered structural properties of language that were claimed to be independent from meaning and context. Whatever the other merits and failings of this approach, this view allows linguists to restrict their attentions to the relatively tractable (if arguably less interesting) domain of form, thereby sidestepping the inherent challenges of semantic representation. In other words, the object of study was shifted to something that could be defined in terms of *some* science — in particular, the nascent fields of formal language theory and computability theory.

In formal grammar terms, a *language* is defined as a set of strings over a given *alphabet* of symbols; a *grammar* describes this language with *production rules* that specify how symbols can be manipulated and combined. The grammar, in other words, provides a mechanism for deciding whether a given string is in the language, corresponding to the binary grammaticality judgments discussed in 1.2.1. Grammars (and their associated languages) can be restricted in the kinds of production rules they allow, and their associated languages accordingly ranked on the Chomsky hierarchy of complexity, increasing from finite cardinality languages (with a finite number of strings) up through finite state (or regular), context-free, context-sensitive, primitive recursive, decidable and recursively enumerable languages. Knowledge of a language, on this view, is identified with knowledge of its generating grammar, and language learning can be formalized as grammatical inference. As in other kinds of inductive inference, a particular grammatical inference problem must specify the hypothesis space (usually some subset of the Chomsky hierarchy); a target language (in the form of its generating grammar); the sample set (how and what type of example strings or other information are made available); a learning algorithm; and some success criterion.

Gold's (1967) seminal investigation into the learnability of various language classes established the paradigm of *language identification in the limit*. Within this framework, the learner is presented with a sequence of example strings labeled according to whether they are in the language (either positive only, or both positive and negative), guessing a grammar after each one. Learning is successful if after a finite time the learner guesses the correct grammar — one that allows the learner to decide which strings are in the target language — and continues to do so thereafter indefinitely. Under these conditions, Gold showed that finite languages are guaranteed to be learnable: although an infinite number of grammars can generate a given finite language, the simple strategy of guessing a grammar consisting of exactly the set of presented examples will succeed in the limit. Infinite languages, on the other hand, cannot be identified in the limit unless negative examples are included, since otherwise nothing prevents the learner from guessing a grammar that generates too large a set. That is, there is no strategy that is guaranteed to converge on the correct grammar in the limit.

Applied to the child's situation, these results bring to light an apparent paradox, the so-called "Logical Problem of Language Acquisition" (Baker 1979). While a definitive formal classification of human languages has proven elusive,[1] for present purposes we need only observe that they have infinite cardinality: given a valid sentence S of English, for instance, an acceptable sentence of the form *S is what she said* can always be generated. Further, caretakers appear not to supply negative examples (Marcus 1993). According to Gold's results, then, children should be theoretically unable to acquire patterns of language more general than those encountered without extending inappropriately beyond those limits — yet developmentally normal children reliably do just that.

Efforts to resolve this conundrum have spurred much productive research on grammatical inference, both as a cognitive challenge and as a mathematical abstraction worthy of study in its own right. While our focus is on the former, a few results from the latter are worth mentioning.[2] In particular, several variations on Gold's theme show how modifications to the learning paradigm's assumptions can yield more promising learning results:

[1]Many syntactic patterns can be represented using context-free grammar rules (also called *phrase structure rules*), as illustrated by the canonical $S \rightarrow NP\ VP$ (capturing the intuition that a sentence S can be generated from a noun phrase NP followed by a verb phrase VP). But it is not clear that context-free grammars are necessary at all levels of description or sufficient for all languages: many morphological patterns are amenable to finite-state characterizations, and a few languages appear to exhibit limited context-sensitivity (Shieber 1985; Culy 1985).

[2]See Pinker (1979) for a review of work through the late 1970s; Angluin & Smith (1983) for a thorough introduction to inductive inference and a review of early theoretical results; Sakakibara (1995) for an extensive survey of work in grammatical inference, covering approaches to learning finite automata, context-free grammars (CFGs), stochastic grammars and non-grammatical representations; and Lee (1996) for a concise summary of work on the induction of context-free languages.

Restricted hypothesis space: Structural constraints on the search space of hypotheses can provide enough bias to make learning tractable without negative evidence. As discussed further in Section 2.1.2, approaches along these lines — particularly those assuming strong innate structural biases — have been especially influential in the child language literature.

More informative input: The learner can infer negative evidence if the sample strings are ordered, such that, for example, shorter strings are guaranteed to appear earlier (Feldman 1972); or if it has access to an oracle that can answer membership queries (Angluin 1988).

Relaxed success criteria: Instead of exact identification, the learner can instead *match in the limit* (guess an equivalent grammar producing an equivalent language), *approach* the target grammar (eventually guess, while rejecting incorrect grammars) (Feldman 1972), or learn the *simplest* grammar consistent with the sample (according to some measure of grammatical complexity, such as that defined in Feldman (1972)). Any primitive recursive language can also be *approximated* to any degree of accuracy using only positive examples, where Wharton (1974) provides criteria for measuring distance between the target and guessed grammar.

The strategies above all reflect more realistic assumptions about the child's learning situation that make the task more tractable, and all appear in some form in the child language acquisition literature, as discussed further in Section 2.1.2.

A more fundamental shift in learning paradigm can be traced to Horning's (1969) work, which shows that the inclusion of input frequencies makes learning more feasible. This approach combines several aspects of the strategies above: the hypothesis space is a stochastic context-free grammar, with probabilities assigned to each production rule; sample data is assumed to be generated by this target grammar, with appropriate associated frequencies; and the learner seeks not an exact (or even an equivalent) grammar, but instead the *optimal* grammar, according to some measure that takes into account both the complexity of the grammar (as mentioned above) and its degree of fit with the observed input data.

Horning's probabilistic, optimization-based framework presaged the rise of Bayesian methods in computational linguistics (and indeed, all walks of artificial intelligence), and it is the direct antecedent of the approach pursued in this work. Of course, the notion of grammar adopted here differs in important ways from these formal grammars: our hypothesis space, sample set and success criteria must incorporate meaning on par with formal structure. Further, while Horning employed an exhaustive search over the space of grammars (enumerating them in increasing complexity), this work will exploit domain-specific search heuristics. But the current work will retain the basic intuition of using an evaluation function that favors simplicity and extend it to more semantically inclusive structures, as discussed further in Section 2.3.4.

2.1.2 Theoretical frameworks

The early developments in grammatical inference just described have had lasting repercussions for both linguistic theory and cognitive science. Gold's results demanded some revision of the problem formulation that would explain how children overcome the apparent mathematical impossibility of the task. Differing tactics for accomplishing this have produced some well-known and persistent theoretical divisions. The most prominent linguistic conflict hinges on whether and to what degree different theories accept formal symbolic grammars as the object of study, and grammatical induction (specifically, identification in the limit) as an appropriate idealization of the task faced by the child. In the acquisition literature, the division is typically framed as a version of the familiar nature-nurture debate, between approaches exploiting different kinds of constraints — genetic or environmental — to stack the deck in the child's favor.

These divisions are, by no coincidence, highly correlated. As noted earlier, the view of grammar most closely associated with Chomsky and his followers, broadly referred to as *Generative Grammar*[3], takes Gold's formalization as not just relevant to but in fact definitive of the child's problem. From this perspective, the best (and only) way to resolve the learnability paradox is to dramatically restrict the hypothesis space, and to posit that these restrictions are innately specified, as suggested by Chomsky's (1965) argument from the poverty of the stimulus. In more recent incarnations of this *nativist* stance, the innate capacity for language is manifested as a *Universal Grammar* with a limited set of language-specific *parameters* whose settings are constrained by universal *principles*. The limited set of options makes the problem formally tractable, since in theory even a small number of input sentences is sufficient to trigger the correct parameter settings for their input language.

Many theorists, however, object to one or more assumptions of the formal learnability approach, and have advocated the inclusion of other sources of knowledge, such as semantic, pragmatic or statistical information, in the domain of study. This group spans several compatible lines of work focusing, variously, on the ways in which language structure reflects its pragmatic and

[3]The term *generative* is polysemous. In its original sense, a *generative* grammar is one that can be used to *generate* the strings of the language; this sense of 'generative' can be taken to apply equally to wide range of grammars. But the term has also come to refer specifically to theories in the Chomskyan tradition (*e.g.*, Transformational Grammar, Government and Binding, the current Minimalist Program). To reduce confusion, I use the capitalized terms *Generative Grammar* or *Generative* when this latter sense is intended, and *syntacto-centric* for the broader class of theories in which syntactic patterns constitute the core object of study. A parallel terminological confusion arises from the terms *formal* and *formalist*, which are frequently associated with Generative Grammar. I avoid the term *formalist*; the usage of the term *formal* reflects the fact that non-Generative grammars, including construction-based grammars, may also be *formal* in a mathematical and computational sense, *i.e.*, specified precisely enough to implement.

communicative functions in context (*functionalism*), grounding in cognitive mechanisms (*cognitive* linguistics), or specific instances of language use (*usage-based* theories). All of these frameworks are thus compatible with theories of acquisition that rely less on innate endowment of linguistic expectations and more on how general cognitive mechanisms, learning abilities and processes of language use may influence language learning (*interactionist* or *emergentist* theories).

While this broad classification does reflect many genuine philosophical oppositions, it also tends to obscure potential common ground across the approaches. As suggested by the proliferation of terms above, the issues involved are multidimensional, and many apparent dichotomies may be more accurately described as reconcilable differences in degree or focus. I will not attempt an exhaustive review here of all of the approaches above and their many points of conflict and overlap.[4] Instead, I summarize the main dimensions of variation that underlie the theoretical tensions above, which serve as orientation for the approach taken here.[5]

Hypothesis space. Theories vary in what they take as the target of learning (rule-based or item-based), whether it includes meaning, how open-ended it is (finite set of parameters or open-ended inventory), and what range of data it must account for (only "core" grammatical structures, or more "peripheral" language, including metaphorical, idiosyncratic or creative language). The key issue is what kind of distinction, if any, is drawn between lexical and grammatical structures. Proponents of Generative Grammar describe knowledge of language as essentially rule-based, with interactions among rules constrained by a finite set of parameters, and transformational or derivational relations among the resulting structures. In such theories, the lexicon is the only open-ended set of structures; it is the only entry point for meaning, and the repository of idiomatic and "peripheral" uses not accommodated by the "core" syntactic patterns of the language. But a variety of *lexicalist* or *monostratal* theories have taken an increasingly unified view of lexical and grammatical knowledge. *Lexical functional grammar* (Bresnan 2001; Dalrymple 2001) and *head-driven phrase structure grammar* (Sag *et al.* 2003; Pollard & Sag 1994), for example, invest lexical items with information traditionally considered the province of grammar, such as the particular argument structure realization patterns licensed by specific verbs. The *construction-based grammar* approach described in Chapter 1 (Goldberg 1995; Kay & Fillmore 1999) represents the limiting case in which there is no hard distinction between lexicon and grammar: linguistic

[4]See Hirsh-Pasek & Golinkoff (1996) for a comprehensive and balanced overview of the developmental literature, and Huck & Goldsmith (1995) for a closer examination of the ideological roots of the divisions in linguistic theory.

[5]My organization follows that of Hirsh-Pasek & Golinkoff (1996) in broad outline. Note that they characterize the two main groups discussed here as "inside-out" and "outside-in" approaches (reflecting their reliance on internal versus external sources of constraints), and compare them in terms of three components: initial structure, mechanism of learning, and source of learning bias. I choose a similar but slightly expanded set of issues for comparison here, in virtue of my interest in the nature of the grammatical representation.

knowledge consists of an open-ended collection of units, ranging from more abstract to more idiomatic structures and encompassing semantic and pragmatic information in units of all sizes.

Prior knowledge. All theories assume the child brings some initial knowledge that guides learning, and that many domain-general skills are present pre-linguistically. The main distinction to draw here is what innate knowledge, if any, can be considered specifically linguistic. Nativists make the strongest assumptions: that there are innate predispositions toward a limited set of syntactic parameters. In some cases (*e.g.*, Pinker (1989)), these are further associated with semantic categories (also universal and innate) by a set of innately specified linking rules. Theories that admit broader categories of information into the target of learning instead posit universals arising mainly from the structure of perception, action and cognition, or from predispositions to notice statistical regularities or indulge in social and goal-oriented behavior (including communication). Linguistic categories, according to this view, need not be innately specified, but rather are *constructed* on the basis of humanly relevant categories; specific packagings of these categories thus exhibit motivated, but not predictable, patterns of crosslinguistic variation (Talmy 1988; Talmy 2000; Lakoff 1987; Langacker 1987) (approaches in this *cognitive linguistic* tradition are discussed further in Section 2.2.2). These privileged notions may act as a universal Basic Child Grammar (Slobin 1985) consisting of prototypical scenes involving a set of co-defined participant roles, such as a manipulative activity scene.

Input evidence. All theories assume that language learning is based in part on perceived phonological forms; typically, for grammar learning, these are simplified and segmented into word strings, on the presumption that children can likewise segment the string by the time word combinations are learned. Most theories also accept the basic generalization that explicit negative evidence, if available at all, is rare and typically unheeded by the child. Theories vary in how much additional information may be included, especially with regard to the meaning or function of the utterance. The classical Gold situation admits no meaning at all; more inclusive theories assume that such utterance forms occur with some semantic interpretation within a broader context of use.

Learning algorithm. The range of learning algorithms employed reflects the restrictions imposed by the other constraints assumed. Since nativist assumptions restrict the faculty of language to vary within the relatively narrow boundaries established by a small number of parameters (generally binary), relatively weak learning mechanisms are necessary; indeed, the child is presumed to *discover* the appropriate parameter settings out of a pre-specified finite set of choices. Proposed mechanisms for this discovery procedure include “bootstrapping” hypotheses that privilege certain cues, either syntactic (Gleitman & Gillette 1995) or semantic (Pinker 1989), as particularly salient or reliable entry points for grammar learning. Theories with a more open-ended hypothesis space tend to rely on domain-general learning strategies, just as in most theories of lexical learning, and to show how linguistic knowledge *emerges* from the interaction and competition of diverse types of constraints (MacWhinney 1998; Bates & MacWhinney 1987; Elman *et al.* 1996). Many theorists have also called attention to the ways in which domain-general learning strategies — such as imitation, bottom-up similarity-based generalization and statistical pattern-matching — may be exploited in lexical and grammatical learning (Bybee 2006; Bybee 1985; Slobin 1985; Maratsos & Chalkley 1980; Tomasello 2003; Clark 1993; Clark 2003).

Many traditional dichotomies in the study of language — between relative emphasis on form or function, innate or environmental constraints, and domain-specific or domain-general mechanisms — make sense in light of the contingent relationships among these components. In particular, the choice of hypothesis space tends to have a determinative effect on the remaining components, with natural affinities between Generative Grammar and nativism on the one hand, and between functionalism and interactionism on the other. Nevertheless, a strict division into two opposing camps does not accurately reflect the range of possibilities above. Many specific theories may make firm commitments along only a subset of the dimensions above. Variants of construction grammar, for example, are all characterized as monostratal, but they are not all committed to cognitively motivated semantic representations or usage-based theories of acquisition or use. Theories also differ in whether or to what degree they consider evidence from the processes of language use, which inevitably involve meaning. Hirsh-Pasek & Golinkoff (1996) thus make a distinction within nativist approaches between *structure-oriented* theories (canonically, that of Chomsky) and *process-oriented* theories (like the bootstrapping theories of Gleitman and Pinker).

The theoretical commitments of the current approach can now be stated in the terms laid out above. As noted earlier, the hypothesis space assumed in this work is a construction grammar in the monostratal tradition, in which meaning is grounded in human conceptualization and embodied experience. The universals assumed here as prior knowledge are primarily domain-general, though a domain-specific predisposition to attend to certain relations between forms, and more generally to relationships between form and meaning, is also included. The input evidence is maximal, in the sense that it encompasses all kinds of experience-driven evidence, including the particular forms and functions of an utterance, the statistics of usage and results of processing. The learning strategy exploits mostly domain-general procedures, though the specific operations used to search the space depend on domain-specific properties of learned constructions.

This theoretical profile is most closely aligned with functionalist and emergentist approaches, but it also takes an intermediate, inclusive stance on some of the dimensions above. Note that it is compatible with the general framework of formal grammar learning, where 'formal' is used here in its computational sense, *i.e.*, referring to grammars that are specified in terms that lend themselves to computational implementation. The particular formal grammar described in Chapter 3 departs from the syntacto-centric tradition in including representations of meaning and context, largely motivated by cognitive considerations. But, assuming the representational challenges involved can be met, there is no inherent conflict between *formal* grammars and meaningful grammars — nor any obstacle to applying statistical learning techniques to meaningful representations, nor to assuming innate predispositions toward certain aspects of form and meaning. Chapter 9 will consider how the approach taken by the current model reconciles some of the tensions above.

2.1.3 The course of acquisition

This section surveys some key developmental milestones of the first two years. As discussed in the last section, the theory of language acquisition to be explored here is maximally inclusive: both genetic and environmental factors contribute to the child's gradual mastery of communication. I structure the discussion around several streams of development, ranging across various aspects of linguistic and extralinguistic knowledge. These streams often intersect, and the specific order and timing along each course is subject to significant individual variation. Nonetheless, some general developmental patterns, along with a core set of sensorimotor, social and linguistic skills, are likely to be present by the time the earliest word combinations are produced (18–24 months).

Sensorimotor and conceptual schemas. Infants inhabit a dynamic world of continuous percepts, and how they process and represent these fluid sensations remains poorly understood. Within the first few months, however, a stable perceptual world emerges from the chaos. By 4–5 months they have developed the expectation that physical objects persist even outside of their immediate perceived environment, and that some objects move on their own while others move as the result of some external cause. Over the next several months, they become familiar with a substantial repertoire of concepts corresponding to people, objects, settings and actions; they also acquire important motor skills like crawling, standing and eventually walking (around 9–12 months). Well before their first word combinations, babies are competent event participants who have accumulated structured knowledge and expectations about the roles involved in different routine events controlled by caretakers and situated in a particular cultural, social and physical context (*e.g.*, meals, baths, play, bedtime, dressing). They know which participants play which roles in an event, with what kinds of results (usually a change of location or some other property of a central participant) (Nelson 1996; Tomasello 1992; Mandler 1992; L. Bloom 1973), and they are sensitive to a variety of spatial and conceptual relationships, such as (in)animacy, causality, agency, containment and support (Mandler 1988; Mandler 1992).

Social and pragmatic skills. Rudiments of social interaction are evident from the first few days of life. Infants are attracted to faces (or face-like configurations) and have a smiling reflex within days after birth. As they gain control over their perceptual and motor abilities, they begin to deploy gestures (such as reaching, by 6–9 months) to achieve their goals. Soon afterward they have the basic skills for directing and inferring the adult's locus of attention, including following pointing gestures and monitoring the adult's gaze and orientation by 9 months, and producing their own pointing gestures by 12 months. Perhaps most impressive are the interpersonal skills youngsters acquire by this time: by their first birthday, children expect people to exhibit goal-oriented, social behavior: people (as opposed to inanimate objects) can affect each other from a distance, and hands (as opposed to sticks) can move in a goal-directed manner. There is also mounting evidence that very young children can infer the intentions of their interlocutors: children as young as 18 months old can imitate *intended* actions of an experimenter, that is, they can successfully complete actions like placing one object on top of another even when the experimenter only *attempts* but does not successfully complete the action (Meltzoff 1995; Tomasello 1995). As discussed below, lexical acquisition appears to be sensitive to these attentional and intentional factors.

Early linguistic forms. It appears that no time is too early to start learning language-specific phonological and intonational patterns: newborns attend preferentially to the language spoken by their mothers, indicating that some learning takes place even while in the womb. This idea of the newborn as a natural statistician has gained additional support from studies demonstrating that 8-month-olds can learn to distinguish high- and low-probability phonological transitions from brief exposure (Saffran *et al.* 1996; Aslin *et al.* 1998). Importantly, some perceptual abilities seem to weaken over time: although newborns can detect arbitrary phonological distinctions drawn from any natural language, within a few months they become specially attuned to those relevant in their native language. On the production front, babies babble at around 6 months and can imitate many phonemes and intonational contours produced by their caretakers around 9 months. They also gain an appreciation for the shape of ongoing discourse, including turn-taking and other aspects of dyadic interaction. By 9 months, children also exhibit some ability to perform morphological segmentation (Jusczyk 1997).

Single words. The first recognizable word forms emerge around 10–14 months. Well before then, however, children are capable of making associations between forms and meanings and thus acquiring some forerunners of bonafide lexical items. Many children produce phonologically consistent forms — reduced or novel sounds that recur in the same situational context or with the same intention — that serve as a transition between babbling and first words. In addition, goal-oriented gestures can function as early communicative signs; indeed, both deaf and hearing children can acquire consistent manual signs as early as 6–8 months, suggesting that the cognitive capacity for making lexical associations is present by that time, and the later timetable for the production of vocal gestures may be attributed to processes of articulatory maturation.

Irrespective of modality, the key development that pushes children firmly into word learning is an appreciation of the *communicative* function of sounds (or gestures), and the gradual shift from associating arbitrary forms and meanings to making more selective mappings based on how other speakers convey their referential intentions. In fact, some one-word utterances (often with characteristic intonational contours) may convey quite sophisticated speech acts; these *holophrases* fall on the cline between single words and telegraphic speech.[6]

Some characteristics of early word learning are reviewed here (see P. Bloom 2000 for a comprehensive overview):

- *Fast mapping*: Lexical items can be acquired based on very few incidental exposures, through a process dubbed *fast mapping* (Carey 1978; P. Bloom 2000). It appears, however, that the conditions under which children can use fast mapping to acquire labels for novel objects become more constrained over time. Very young children (12-13 months) easily acquire labels for objects in their focus of attention; after only a few exposures, they can later correctly pick the object out from a line-up of candidate objects in response to the new label. Experiments have shown that a variety of "forms" — including gestures, artificial sounds and pictograms (Namy 2001; Woodward & Hoyne 1999) — can successfully serve as a label for a novel object. But older infants (around 20 months) appear to be *less* successful at fast mapping under

[6] The term *holophrase* has also been applied to fixed expressions that are functionally indistinguishable from single words but convey a complete speech act. That is, the child may perceive a multi-word combination as a single unit and treat it as such until it is reanalyzed in terms of its constituent parts.

those conditions, suggesting that they may have narrowed the realm of possible referential forms to those generated by the human articulatory tract (or manual system in the case of sign language). They are also sensitive to the adult's attentional and intentional state: 15-month-olds preferentially acquire labels if joint attention on the object has been established (Baldwin 1993). Most strikingly, successful mapping may also be contingent on the perceived referential intention of the speaker. Tomasello (1995) describes an ingenious experiment in which experimenters announce they are looking for a particular named object concealed in one of several buckets. As they retrieve and display objects from the buckets, they evince either delight or dismay, presumably conveying, respectively, a successful or failed search for the named object. Both two-year-olds and 18-months-olds learn labels for objects contingent on successful searches.

- *Non-ostensive learning*: Some early words may be learned in the stereotypical ostensive situation exemplified by a mother pointing to the family pet while saying "cat". But while straightforward associative process can account for object labels and proper names, some of the most frequent words in children's early speech do not have physically available referents and thus cannot have been learned based on temporal coincidence of a sound with a stable perceptual experience. These include *function words* (*e.g.*, English *no*, *uh-oh* and *bye-bye*) that depend on notions of social interaction and the achievement (or failure) of goals (Gopnik 1982), as well as verbs or other *relational terms* (L. Bloom 1973) (*e.g.*, *up*, *go* and *more*) that refer to transient actions or events. Furthermore, cultures vary widely in the accessibility of ostensive learning situations, and some (*e.g.* Kaluli) appear not to provide any explicit ostension.

- *Generalization*: Children extend the labels they acquire in a contextually constrained but nevertheless productive way. Children appear to avoid the logical pitfalls of inductive inference identified by Quine's (1960) discussion of the infinite possible referents for the word *gavagai* uttered as a rabbit scampers by: most of their generalizations are appropriate, and even their overextensions are "typically reasonable ones, honest mistakes" (P. Bloom 2000:38). Most confusions have a discernible basis in similarity, for example of shape (*ball* referring to the moon) or event-type (using *up* and *down* interchangeably, as in L. Bloom (1973); or for different aspects of similar events (using *fly* to refer to both birds and the action of flying).

In short, word learning is considerably more complex than forming simple associations. Words can be learned from few exposures, without negative feedback, without a stable or tangible referent, and in spatially and temporally non-contiguous situations.

Researchers have posited a variety of constraints and principles to account for these facts of lexical acquisition. These include domain-specific biases, for example to pay special attention to shape (Landau *et al.* 1988) or prefer labeling whole objects (Markman 1989), as well as more general pragmatic principles like Clark's (1993) principles of contrast (different forms map to different meanings) and conventionality (particular conventional forms for particular meanings preferred). P. Bloom 2000 argues for theory of mind as the dominant factor in word learning: children infer lexical mappings that reflect and encode referential intent on the part of their interlocutors. That is, they understand that adults use conventionalized forms to direct attention to aspects of the environment or accomplish other communicative goals.[7] This account may be considered an updated version of Augustine's evocative description of language learning as relying on aspects of social and pragmatic intelligence, such as

> the motion of their body, the natural language, as it were, of all nations, expressed by the countenance, glances of the eye, gestures of the limbs, and tones of the voice, indicating the affections of the mind, as it pursues, possesses, rejects, or shuns. (61)

The current work will not model all the sophisticated interpersonal skills that appear to be involved in lexical acquisition. But these skills presumably remain available as children begin to learn larger and more complex units of language; the model thus assumes that both the input the learning and the mechanisms of language comprehension approximate some of these precocious social skills.

Word combinations and early syntax. Most children spend several months in the single-word stage before the first word combinations appear, around 18–24 months. Overt grammatical markers (inflectional morphology and function words) and more complex productive utterances emerge around 24–36 months. These time estimates refer to production data; although children can often respond appropriately to multi-word utterances even while in the single-word stage, it is generally difficult to discern how much of their behavior is driven by their (considerable) pragmatic skills and how much can be attributed to the comprehension of linguistic devices.[8]

Some examples of early English word combinations include *sit down*, *throw off*, *want that*, *Daddy play* and *close door* (Sachs 1983); they tend to be telegraphic in nature, typically lacking many closed-class gramamtical morphemes like determiners, verbal auxiliaries and verbal inflections. As might be expected from the discussion in Section 2.1.2, the nature of any underlying representations for these combinations has been subject to much debate. Particular attention has focused on whether

[7]This point of view fits well with Tomasello's (1999) hypothesis that non-human primates, despite their prodigious cognitive and social abilities, do not match the natural language learning abilities of human children because they lack the capacity for (or at least the predisposition toward) representing, reasoning about and manipulating the attentional and intentional states of others.

[8]There is some experimental evidence that infants in the single-word stage (17.5 months) can use word order as a cue to meaning (Hirsh-Pasek & Golinkoff 1996), though this may be only in conjunction with semantic cues.

early utterances are best analyzed as patterns based on syntactic categories or relations, semantic categories or relations, or distributional facts about specific lexical items.

The evidence suggests that all three factors play some role, though they may feature more or less prominently at different stages of acquisition. Semantic and distributional factors seem to play a dominant role for the earliest word combinations, as evidenced by crosslinguistic recurrence of certain broad categories of semantic relations, such as *agent* + *action*, *action* + *object* and *attribute* + *object* (Brown 1973). More specific semantic properties — such as size and color rather than attribution in general, or animacy rather than agenthood — have also been proposed (Braine 1976). In the extreme, many early combinations need not involve categories at all but rather instantiate patterns with specific words in fixed positions (Bowerman 1976; Braine 1976; Maratsos & Chalkley 1980), as suggested by Braine's (1963) proposed *pivot grammar*. More recent studies of individual courses of acquisition also support the idea that the earliest constructions are *item-specific* (Tomasello 1992; Tomasello 2003), and more generally that the acquisition of multi-word constructions bears many of the characteristics of lexical acquisition mentioned above. (Further evidence for these ideas is presented in Section 2.2.3, with related learning proposals in Chapter 6.)

As children move into later stages, they produce longer sentences containing more closed-class morphemes traditionally considered grammatical markers. While the underlying representation remains subject to debate, later productions exhibit much more variability. They also make segmentation and production errors that suggest they have acquired at least partially abstract patterns, such as the *X-er* pattern evidenced in (2–1a) or the overgeneralized version of the RESULTATIVE construction in (2–1b):

(2–1) a. Daddy, do you need to ham something with your hammer? *(Ariel, 2;11.12)*
b. Let me cold it for you. *(Ari, 2.9.1, in play kitchen)*

These instances reflect the more general observation that even overproductions tend to be semantically well-motivated and intepretable in context, employing adult-like syntactic structures to express complex communicative intentions: by the end of their third year, they are well on their way to mastering the forms, meanings and functions of their native languages.

* * *

This section has given a high-level overview of the theoretical proposals and empirical findings that bear on the study of language acquisition; many other distinctions and issues remain unexplored here. The most relevant points are summarized as follows:

- The formal grammar induction paradigm launched by Gold, though an unsatisfactory idealization of the child's task, can be modified to make more realistic assumptions. The current model will pursue the optimization-based line of inquiry begun with the inclusion of frequency information in Horning's work and extend it to include representations of meaning in the hypothesis space, input and success criteria.
- The theory of language to be adopted here is formal (in the computational sense), but its assumptions have more in common with theoretical outlooks that have not traditionally been associated with formal representations: construction grammar, cognitive linguistics and emergentist, usage-based approaches to learning.
- By the stage of learning addressed by the current model, the child has access to a battery of sensorimotor skills, considerable social-cultural acumen, and both statistical learning and fast mapping abilities. Although many formulations of the language learning problem discount these resources or limit them to lexical acquisition, they will be exploited in the work developed here as rich sources of information for the learner.

2.2 Foundations

We now turn to the more specific assumptions underlying the approach taken here: (1) the target of learning is a construction, or a pairing of form and meaning; (2) meaning is embodied; and (3) language learning is usage-based. These ideas have been regularly applied with little controversy to lexical items; this section highlights evidence supporting a similar view of multi-unit expressions.

2.2.1 Constructions

The basic unit of linguistic knowledge is taken to be a pairing of form and meaning, or a *construction*. This insight is shared by a family of proposals collectively referred to as *construction-based grammars* (Kay & Fillmore 1999; Lakoff 1987; Langacker 1987; Goldberg 1995; Croft 2001). The constructional view of grammar is summarized by Goldberg & Jackendoff (2004) as follows:

> a. There is a cline of grammatical phenomena from the totally general to the totally idiosyncratic.
> b. Everything on this cline is to be stated in a common format, from the most particular, such as individual words, to the most general, such as principles for verb position, with many subregularities in between. That is, there is no principled divide between 'lexicon' and 'rules'.
> c. At the level of phrasal syntax, pieces of syntax connected to meaning in a conventionalized and partially idiosyncratic way are captured by CONSTRUCTIONS. (532)

For the current model, the crucial theoretical commitment in construction-based approaches is to linguistic representations that are symbolic, unified gestalts. I discuss each of these aspects in turn.

Constructions involve a **symbolic** relationship between form and meaning, in a sense consonant with de Saussure's (1916) notion of a *sign*, composed of two parts: *signifier* (a materially produced representation, such as the word form "dog") and *signified* (a concept, such as the animal category dog), corresponding roughly to form and meaning, respectively.[9] Constructions, like signs, are (to a large extent) arbitrary,[10] in that the sound "dog" has no motivated connection to the concept of dog; conventional, in that the relationship is valid only in the context of a community that accepts it as such (*e.g.*, the community of English speakers); and intentional, in that they are used with referential or communicative intent.[11] Indeed, lexical constructions (*i.e.*, words) are the canonical linguistic sign. The constructional view assumes further that *all* linguistic units can be similarly described as based on a *symbolization* relationship, to use Langacker's (1991) term. That is, the CAUSED MOTION construction cited in Chapter 1 as the basis for sentences like *Mary pushed the napkin off the table*, for example, involves an arbitrary, unpredictable relationship between aspects of form (word order) and aspects of meaning (the relevant scene of caused motion); it is conventional for speakers of English; and it is used intentionally by a speaker to predicate the relevant event.

Constructions of all kinds can be captured in a single **unified** representation. This is the sense indicated by (b) above, and discussed in Section 2.1.2 as the monostratal extreme of lexicalist grammatical frameworks. This representation should encompass expressions of all sizes (from morphemes and words to larger phrasal and clausal units with internal structure), at all levels of abstraction (from frozen idiomatic expressions to more general linguistic principles), and both open-class (or *content*) words and closed-class (or *function*) words (*e.g.*, conjunctions, prepositions and determiners). All of these are assumed to share an underlying symbolic relationship between form and meaning.

Constructions function as **gestalts** — that is, each construction is a whole whose behavior is not determined by the behavior of its constituent parts; they thus exemplify the phenomenon studied by the Gestalt school of psychology. Here 'behavior' includes both form and meaning, though I focus on the latter as the more salient departure from standard approaches to meaning in gram-

[9]In contrast to the structuralist dyadic relationship, signs in the semiotic tradition founded by Charles Sanders Peirce are triadic, with two components roughly corresponding to the signifier and signified, and a third for the real-world object or referent of the sign (*e.g.*, a specific dog it references) (Peirce 1976). As discussed in Chapter 5, the current model has an analogue to this third semiotic component in the form of the *resolved referent*.

[10]See Hinton *et al.* (1996) for discussions of iconicity in language; other salient counterexamples include onomatopoeia, phonaesthemes (Bergen 2004) and the motivated semantic functions of reduplication (Regier 1998).

[11]Saussure construed the word 'symbol' as encompassing more iconic relationships and thus preferred to use 'sign' for fully arbitrary linguistic relationships. I will gloss over this distinction here, since it is not crucial for the current discussion, nor clear-cut in its application to all linguistic signs (see footnote 10).

matical theory. Construction grammarians reject the traditional premise that sentences (and other structured linguistic units) derive their meaning as a strictly compositional function of their constituents, bottoming out with words or morphemes and their associated meanings. Rather, a syntactic pattern itself may also contribute a particular conceptual framing, as noted in claim (c) above.

Evidence for the constructional view has come primarily from linguistic investigations into the relationship between patterns of form and meaning, using grammaticality and interpretability judgments like those discussed in Section 1.2.1. Some of the earliest studies in the constructional literature focused on partially idiomatic expressions, such as the LET ALONE construction (Fillmore *et al.* 1988) shown in (2–2). Such constructions have conventionalized, idiomatic interpretations, but they are more flexible than frozen idioms like *by and large* and *kick the bucket*, since they include both fixed and variable elements. They thus occupy an intermediate position on the cline of idiosyncrasy mentioned in (b) above. Crucially, unlike frozen idioms, which are often treated as "big words" — *i.e.*, *kick the bucket* can be codified as a special lexical entry with the same meaning as one sense of the verb *die* — partial idioms must be interpreted with respect to their variable arguments, and subject to the specific syntactic and semantic constraints they impose. For example, the LET ALONE construction can take a range of expressions for its two variable constituents, as illustrated by (2–2a) and (2–2b), but they are required to be of compatible syntactic and semantic types and to have meanings that can be construed as comparable along some inferred dimension (conditions not satisfied by (2–2c) and (2–2d)).

(2–2) LET ALONE (Fillmore *et al.* 1988)
  a. Harry couldn't smile, *let alone* laugh.
  b. Harry couldn't afford a used bike, *let alone* a new car.
  c. *Harry couldn't smile, *let alone* a new car.
  d. *Harry couldn't afford a new car, *let alone* a used bike.

Other constructions studied include the WHAT'S X DOING Y? construction (Kay & Fillmore 1999), as in *What's that blue thing doing here?*; the comparative correlative (Culicover & Jackendoff 1999; Michaelis 1994), as in *the more the merrier* and *The bigger they are, the harder they fall*; and the DEICTIC LOCATIVE or THERE-construction (Lakoff 1987), as in *There goes Harry with a red shirt on* or *Here comes the sun*. Note that this last case has no fixed lexical material; it permits a range of motivated variations. Like the others, however, it can still be described as exhibiting variability in its permitted arguments and idiosyncrasy in its syntactic, semantic and pragmatic constraints. All of these examples suggest that constructional categories are, like other conceptual categories, best described not as classical formal categories (*i.e.*, defined by a set of necessary and sufficient conditions) but as *radial categories* (Lakoff 1987) with prototype structure and graded membership.

A different kind of evidence for the constructional view comes from the semantic constraints associated with even more general syntactic patterns, such as the CAUSED MOTION construction discussed already. The DOUBLE OBJECT or DITRANSITIVE construction (Goldberg 1995) in (2–3) is another such *argument structure construction*, in this case pairing ditransitive syntax with a transfer scene in which a sender or source entity transfers some item to a recipient. On this account, it is the construction-level scene that imposes the benefactive reading in (2–3b) (albeit with a modified sense of creation with *intent* to transfer), and the implicit need for an agentive, intentional recipient that renders (2–3c) anomalous.

(2–3) a. Harry kicked Susan the ball. *(transfer scene)*
b. Harry baked Susan a cake. *(benefactive transfer scene)*
c. *Harry kicked the door the ball.

Similar observations have been made for the RESULTATIVE construction (Boas 2003; Goldberg & Jackendoff 2004). All of these examples, along with the more idiosyncratic examples discussed above, demonstrate the key constructional properties identified here: each can be seen as a symbolic gestalt that pairs form and meaning in a conventionalized way, captured within a uniform framework that encompasses constructions of all sizes and levels of abstraction.

Finally, some support for the constructional view of grammar comes from psycholinguistic studies (developmental evidence is deferred until the discussion of usage-based learning in Section 2.2.3). Kaschak & Glenberg (2000) describe an experiment in which adult subjects interpret sentences using novel denominal verbs, such as *Lyn crutched Tom her apple so he wouldn't starve* (based on the double object construction) or *Lyn crutched her apple to Tom so he wouldn't starve* (based on the caused motion construction). Subjects overwhelming inferred a transfer event for sentences using the double object construction, compared with a simple transitive event (something acting on something else) for the caused motion construction. These semantic associations could not be attributed to properties of the (novel) verb alone, nor to the arguments (which were constant over the two constructional conditions). They appear, rather, to stem from the syntactic form itself, together with the general denominal affordances of English, as predicted by the constructional view.

2.2.2 Embodiment

Any venture into the domain of meaning invites skepticism on many grounds. Diverse thinkers of considerable repute have championed (and disparaged) a range of positions on the ontological status of meaning and how it interacts with language learning and use; the matter will by no means be laid to rest here. Nevertheless, the approach explored in this work commits us to the proposition that it is possible, in principle, to establish a scientific basis for (many aspects of) meaning, and necessary, in practice, to do so in a computationally explicit manner. The working assumption adopted here, as stated earlier, is that meaning is *embodied*: it is grounded in the interaction between the human nervous system and its physical and social context. This includes all factors that may affect human conceptualization, from features of action and perception to processing constraints on attention and memory. I discuss two broad categories of evidence supporting this embodied basis of meaning: (1) crosslinguistic patterns of embodiment in language; and (2) psycholinguistic and biological studies of language processing. I also take the evidence as consistent with, and in some cases directly supportive of, the stronger claim about the nature of language understanding expressed by the Simulation Hypothesis in Chapter 1.

Crosslinguistic conceptualization

Crosslinguistic studies suggest that linguistic structures of all kinds evoke meanings that are motivated by, though not strictly predictable from, features of the human neurophysiology. Berlin & Kay's (1969) landmark study found that the meanings associated with basic-level color terms across languages exhibit prototype structure with graded membership, and that the statistically most representative examples coincide with focal colors — *i.e.*, those that elicit peak response wavelengths of retinal color cones. More recent results, reviewed by Kay & Regier (2006), confirm that color naming is subject to universal tendencies, though the particular foci and boundaries that arise in natural languages may result from a range of psycho-physical, environmental and population factors affecting color perception: "nature proposes and nurture disposes" (Kay & Regier 2006:53).

It appears, then, that color concepts are neither abstract ideals (in the Platonic sense) nor arbitrary subsets of the color spectrum. Similar claims may apply to other open-class content terms associated with concrete perceptual and motor domains (such as object labels or motor actions). But what about domains associated with closed-class function words (like prepositions and conjunctions) and other grammatical markers (*e.g.*, morphologically marked case)? Cognitively motivated approaches to semantics suggest that even these domains draw on aspects of embodied experience (such as static physical configurations like containment, contact, support and proximity; dynamic events such as motion along a path and force transmission; temporal relations and event structure; intentional structure and goal achievement; and asymmetries in perceptual salience) — and further, that it is precisely these *grammaticizable* notions that are the best candidates crosslinguistically for closed-class marking (Talmy 2000).

Terminological distinctions in the cognitive linguistics literature reflect the high correlation among some of these conceptual domains. Both Langacker's (1991) *trajector-landmark* and Talmy's *figure-ground* distinction refer to asymmetric attentional relationships in which the orientation, location, or motion of one entity (the *trajector* or *figure*) is defined relative to another (the *landmark* or *ground*); these roles are specialized as the *agonist* and *antagonist* in the realm of *force dynamics* (Talmy 1988). More generally, a number of *image schemas* (Johnson 1987; Lakoff & Johnson 1980) have been proposed to capture recurrent patterns of sensorimotor experience. We will discuss the most relevant of these further in Chapter 3.

As in the color case, specific terms exhibit prototype structure and may overlap in their extensions, and specific languages differ in precisely how they carve up the same (continuous) conceptual space. The containment relationship expressed by English *in*, for example, conflates tight-fitting and loose-fitting containment, which are expressed by two distinct markers in Korean (Choi & Bowerman 1991). But the patterns of variation observed appear to have an embodied basis. Slobin (1997) discusses a study by Schlesinger (1979) documenting a conceptual continuum of meanings associated with the preposition English *with*, ranging from comitative (*Jason went to the park with Theo*) to instrumental (*Phoebe eats with chopsticks*). The study shows how twelve different languages divide the continuum between different markings at different points; a similar range seems to hold between containment and support, which are conflated in Spanish *en* (and divided in English, for example, between *in* and *on*). The presence of this continuum suggests that the distinctions made by different languages are neither arbitrary nor genetically determined. Instead, they are shaped by the cognitive apparatus common to users of all of these languages, and mediated by the accidental and opportunistic forces of language development.[12]

Evidence from language understanding

The claim that meaning is embodied extends beyond the conceptual categories ascribed by linguists to individual constructions: the current work is also concerned with how constructions dynamically combine in context to give rise to a particular set of inferences. Although mechanisms of language use remain poorly understood, a number of cognitive psychologists have stressed the importance of perceptual representations (Barsalou 1999), affordances (Glenberg & Robertson 1999) and perspective-taking (MacWhinney 2005) in language. These suggestions are consistent with the idea that language may exploit the same structures used in action, perception and other neurally

[12]See Hopper & Traugott (1993) for diachronic evidence of a cline of grammaticization between lexical items and grammatical markers, including many examples of embodied lexical items that gradually take on more grammatical form and functions, *e.g.*, English noun *back* (referring to a body part) leading to the prepositional expression *in back of* (behind) and eventually to a particle with both spatial and temporal senses (as in *go back* or *think back*).

grounded activities, and that patterns of inference may be understood as involving simulative imagination based on those structures. Below I examine some recent behavioral and neurobiological studies that lend support to these ideas.

Several psycholinguistic experiments offer behavioral evidence for the automatic and unconscious use of perceptual and motor systems during language processing. Some of these show how incompatibilities between actions the subject performs and language the subject hears can influence processing time: Subjects processing sentences encoding upward motion (*e.g.*, *The ant climbed*) take longer to perform a visual categorization task in the upper part of their visual field (Richardson *et al.* 2003), and subjects performing a physical action in response to a sentence take longer to perform the action if it is incompatible with the motor actions described in the sentence (Glenberg & Kaschak 2002). A few studies offer more direct evidence that language may involve mental simulation. Subjects take longer to make decisions about fictive motion sentences (*e.g.*, *The highway runs through the valley*) given story contexts involving faster motion, shorter distances and less cluttered terrains (Matlock 2003). Comprehension of sentences based on the double object construction with novel denominal verbs (like those mentioned in Section 2.2.1, *e.g.*, *Rachel chaired the scientist his mail* depends on whether the affordances implied by the story context (*e.g.*, the presence or absence of a suitably wheeled chair) supports the semantic constraints of a transfer scene (Kaschak & Glenberg 2000). Note that this experiment suggests that the semantic contribution of the argument structure construction is not in itself enough to license a particular interpretation. Rather, the overall interpretation depends on features of the entire scene (as suggested by the boulder-sneezing example in Section 1.2.1). As expressed by Kaschak and Glenberg's Indexical Hypothesis (2000): "Meaning arises from the mesh of affordances guided by intrinsic biological and physical constraints and the scene or goal specified by the construction."

Neurobiological evidence centers on experiments from the study of *mirror neurons* (Gallese *et al.* 1996; Rizzolatti *et al.* 1996), which fire during both recognition and execution of specific actions. Gallese & Lakoff (2005) present a detailed argument for how mirror neurons may serve as the basis for embodied concepts and the Simulation Hypothesis. (See Gallese & Lakoff (2005) for further references, and Section 9.3.4 for additional discussion.) Most relevantly, a growing body of evidence indicates that areas of motor and pre-motor cortex associated with specific body parts are activated in response to motor language referring to those body parts. Verbs associated with different effectors (*e.g.*, *chew*, *kick* and *grab* for the mouth, leg and hand, respectively) display more activation for the appropriate associated regions of motor cortex (Pulvermüller *et al.* 2002; Hauk *et al.* 2004). Passive listening to sentences describing mouth, leg and hand motions also activates corresponding parts of pre-motor cortex (Tettamanti *et al.* 2005). These experiments provide suggestive evidence for an integrated, multimodal action representation that serves as a common substrate for action, perception and language.

2.2.3 Usage

The current work assumes that language structure emerges from language use. This claim is associated most directly with *usage-based* theories of language (Bybee 1985; Langacker 1987) and, more recently, usage-based theories of acquisition in the developmental literature (Tomasello 2003; Clark 2003). All of these share a commitment to the idea that linguistic knowledge is the totality of structures acquired through bottom-up, data-driven processes over the learner's history of usage. The resulting structured inventory may vary in complexity, abstractness and degree of entrenchment. Usage-based models are thus fully compatible with the constructional view discussed above. Note, however, that *usage* has multiple related senses; some of these are also closely affiliated with other proposals in the literature:

- *Instances of use*: Individual instances of use serve as *exemplars* that are memorized and then extended to other situations via analogical reasoning or generalization. This idea has been explored for phonological learning in the form domain, as well as for cross-domain mappings (e.g., words and their meanings; utterances and their accompanying situations).

- *Usage patterns*: Structure emerges from long-term *distributional* patterns of use. This idea is consistent with the evidence noted earlier that children are sensitive to statistical patterns in phonological data (Saffran *et al.* 1996); a similar approach may be used to discern statistical correlations between form and meaning (Maratsos & Chalkley 1980). More recently, Gahl & Garnsey (2004, 2006) have shown that speaker pronunciation may reflect syntactic and distributional probabilities, blurring traditional distinctions between grammar and usage.

- *Functions of use*: Linguistic units are used in particular contexts by speakers with specific communicative intentions and effects. That is, linguistic knowledge is seen as including patterns of form, meaning and function (as assumed by functionalist and construction-based approaches to grammar). Aspects of pragmatic function may also be salient to the child and thus a reliable learning bias (Clark 1993; Budwig 1995; Bates 1976).

- *Processes of use*: Linguistic units should facilitate the processes of language use. Thus, constructions that are easy to recognize or produce, recur with high frequency, or have more predictive power should be learned most easily. They may also be directly prompted to bridge gaps in the learner's incomplete grammar and thus reduce errors or uncertainty.

These senses are compatible and mutually reinforcing: individual instances consist of linguistic forms used for particular functions in context, and grammatical structure emerges based on those that are statistically most helpful for the processes of language usage. All of these are crucial to the current model. I present some evidence supporting these usage-based assumptions, as well as some usage-based proposals most relevant to the learning problem at hand.

Evidence

Developmental studies suggest that the earliest constructions are tightly linked to particular instances of usage. (Tomasello 1992) observed that his daughter's early utterances consisted largely of verb-specific patterns used for particular kinds of action-specific scenes. These *verb island* constructions — so called because each forms an independent island of organization — appear to develop along separate trajectories, with more general patterns emerging much later. Additional studies by Pine & Lieven (1993) and Lieven *et al.* (1997) have found that many early utterances can be characterized as distributional patterns based on specific lexical items (Tomasello 2003; Israel To appear).

Besides diary-based analyses, Tomasello and his colleagues have compiled extensive experimental evidence in support of lexically specific, or *item-based*, learning (Tomasello 2003). Elicitation tasks show that most two-year-olds produce sentences using nonce verbs in the same syntactic patterns modeled for them by adults, even when the situational context is biased toward a transitive scene. For example, verbs modeled using intransitive syntax (Tomasello & Brooks 1998), passive syntax (Brooks & Tomasello 1999) and even with "weird word order" (*e.g.*, *The cow the horse is meeking* is in SOV order, a logical possibility exploited by many languages but not present in English) (Akhtar 1999; Abbot-Smith *et al.* 2001) were used in the modeled form. Children were able to generalize to new nominal arguments, ruling out a fully imitative strategy, and the percentage of children regularizing to transitive syntax increased with their age (Tomasello 2000), consistent with previous studies showing that children around 4 and 5 years old reliably generalize novel verbs when semantically biased (Pinker *et al.* 1987).

Proposals for learning

Much work in the developmental literature identifies the kinds of information that may play a role in language learning and corroborates the general principles of usage-based learning expressed above. By comparison, relatively few theorists have directly addressed the actual *processes* by which new linguistic mappings are acquired. This section summarizes some notable exceptions that serve as direct antecedents of the learning strategies to be adopted in the current model. (See Chapter 6 for further discussion.)

An early and extensive set of concrete proposals appear in Slobin's (1985)'s catalogue of potential Operating Principles for children acquiring form-meaning mappings.[13] The principles are constructive; they do not make strong assumptions about any fixed set of parameters being rigidly set by the child. Rather, they take the form of directives to pay attention to specific kinds of information (*e.g.*, co-occurrence; frequency; correlation of meanings with forms) and aspects of utterances (*e.g.*, variation, similarity, salient portions of the utterance form, word order). They also include what might be considered "meta-strategies" for noticing the success or failure of the current set of principles and altering them accordingly. The diversity of heuristics proposed can be seen as staking out a centrist position that acknowledges the role of both domain-specific biases and domain-general learning principles. Moreover, the explicit advocacy of frequency as the basis for learning anticipates current trends toward statistical learning. While not all of the proposed operating principles play a role in the model described here, the overall approach of combining usage-based heuristics with statistical learning principles serves as a blueprint for the optimization learning strategy to be employed in this work.

The growing body of evidence in support of item-based learning serves as another key constraint on the learning model. Tomasello's (1992) original Verb Island Hypothesis — that phrasal and clausal constructions are first learned on a verb-specific basis, and only later generalized to form the canonical transitive and intransitive constructions — has in the interim been extended on a much broader basis to instance-based learning. Most work in this domain has focused on documenting the kinds of learning and generalization that do (or do not) take place at specific ages, but a few proposals attempt to explain these phenomena. The compression of similar schemas may lead to more abstract constructions. Existing item-based constructions also serve as a kind of base case that children may adapt to new situations with minimal adjustments, for example by adding or dropping arguments to express or elide specific scene participants (Tomasello 2003).

[13] Peters (1985) proposes an analogous set of principles for perception and segmentation. Though less directly relevant to the current work, these demonstrate the continuity in the overall approach across different levels of linguistic analysis.

Comparatively less work has addressed how aspects of language processing may favor the acquisition of certain constructions over others. The most concrete proposals in this realm come from Clark's work on lexical acquisition, which emphasizes the importance of pragmatic principles that bias the child toward constructions that are easy to learn, recognize and produce (Clark 1993). Specifically, Clark (1993; 2003) proposes that acquisition of morphologically complex words may incorporate biases toward simple forms and transparent meanings (*i.e.*, words with meanings that are based on the meanings of known subparts). These ideas are easily extended to phrasal and clausal constructions, and in particular resonate with the dual biases toward simplicity (with respect to grammar representation) and usefulness (with respect to a model of comprehension and the data encountered) to be quantified by the model's evaluation strategies.

2.3 Formalisms

Computational approaches to language are, like their linguistic and psychological counterparts, divided into several frameworks that make different representational and algorithmic assumptions. Historical ties between language and logic, and more recently between the study of formal grammars and syntacto-centric linguistic theories, led to the development of early natural language processing systems in which a formal grammar (typically a context-free grammar) provided the syntactic core, and some variant of first-order logic served as the basis for semantic representation. These systems, however, have proved too brittle to extend beyond limited application domains. Moreover, since language is taken to be "AI-complete" (in the sense that, in the limit, it relies on all the fields of artificial intelligence, including vision, speech, knowledge representation, planning and inference), logically based approaches to language dominant in the 1980s have proved susceptible to the well-known pitfalls of inference (the frame problem), uncertainty (ambiguity) and robustness under dynamically changing conditions.

During the 1990s, the success of statistical methods in speech recognition and other areas of artificial intelligence spread gradually into computational linguistics. In particular, tasks like part-of-speech tagging and syntactic parsing are now nearly universally approached as probabilistic inference based on large corpora. Semantically and pragmatically oriented tasks, however, have largely lagged behind the statistical revolution. This lag stems in part from the dearth of appropriate semantically rich data, as well as from the contested status of meaning in linguistic theory.

The preeminence of meaning in the current model, as well as the cognitively oriented constraints on memory and processing, prevent many of these mainstream approaches from being directly applicable. Nonetheless, several ideas from the logical and statistical traditions have had a strong influence on the current model, as well as its direct antecedents in the NTL project. This section briefly surveys the main concepts and tools needed for formalizing the foundational assumptions of the current model, focusing on its most relevant forerunners; broader connections to these and other related approaches will be discussed in Chapter 9.

2.3.1 Unification-based grammar

The constructional assumptions of the model are most compatible with those of *constraint-based* or *unification-based* grammatical frameworks, such as that described by Shieber (1986) and used in Head-Driven Phrase Structure Grammar (Sag *et al.* 2003). Such approaches represent linguistic units, also called *types* or *signs* (in the Saussurean sense mentioned in Section 2.2.1), as bundles of *features*, where pairs of features and values capture associated properties like the familiar gender and number. These feature bundles have been formalized as sets of feature-value pairs called *feature structures*. Feature structures may have complex structure, since a feature's value may itself be a feature structure. Formally, they are directed graphs whose edges correspond to features and nodes correspond to values.

Features and feature-based representations have long been used to capture linguistic generalizations, even in rule-based grammars. For example, attribute grammars associate symbols in a standard context-free grammar with feature structures and allow grammar rules to assert constraints over these. In constraint-based grammars, feature structures can serve as a locus for various kinds of information, including phonological, orthographic, syntactic and semantic constraints. The information content of two feature structures can be combined, or *unified*, if their features and values are compatible. This *unification* operation is particularly well-suited for capturing the ways in which multiple linguistic units can contribute different kinds of information to a composite structure. Each word in the phrase *the big dog*, for example, contributes to the whole in different but compatible ways, as crudely illustrated by the simple feature structures in Figure 2.1. Unification may be prohibited when these contributions clash; the unacceptability of **a big dogs*, for example, can be analyzed as resulting from clashing values for the number feature on *a* and *dogs*.

Most unification-based grammars exploit additional devices to capture structured relationships among linguistic units. Feature structures may be associated with a type in an inheritance hierarchy that determines which features are applicable and further constrains unification to require compat-

$$\begin{bmatrix} \textit{the} & \\ \text{category :} & \\ \text{size :} & \\ \text{number :} & \\ \text{definite :} & \text{true} \end{bmatrix} \sqcup \begin{bmatrix} \textit{big} & \\ \text{category :} & \\ \text{size :} & \text{big} \\ \text{number :} & \\ \text{definite :} & \end{bmatrix} \sqcup \begin{bmatrix} \textit{dog} & \\ \text{category :} & \text{dog} \\ \text{size :} & \\ \text{number :} & \text{singular} \\ \text{definite :} & \end{bmatrix} = \begin{bmatrix} \textit{the big dog} & \\ \text{category :} & \text{dog} \\ \text{size :} & \text{big} \\ \text{number :} & \text{singular} \\ \text{definite :} & \text{true} \end{bmatrix}$$

Figure 2.1. An example of unification: feature structures with compatible role values can be unified into a single structure.

ible types on all unified structures. Additional extensions allow multiple inheritance and default values; see Jurafsky & Martin (2000) for an overview and further references.

Unification-based grammars are in many respects a natural fit for representing the constructional pairings of our present concern. In practice, however, many existing formalisms inherit theoretical baggage that make them less compatible with the constructional assumptions laid out in Section 2.2.1, for example by employing meta-principles of combination that privilege form over meaning, assuming strictly compositional semantics, or treating lexical and grammatical units as theoretically distinct. Thus, while the formalism adopted in this work has at its core a unification-based representation, it is also designed specifically to satisfy these construction-based constraints, along with those of the broader simulation-based model of language understanding.

2.3.2 Extending parsing to constructions

The problems of identifying syntactic and semantic structure have traditionally been segregated under the respective rubrics of syntactic parsing and semantic interpretation. The current construction-based approach requires a tighter integration of these tasks, in which constraints from both form and meaning together determine both the underlying constituent structure of an utterance and its corresponding semantic interpretation. While the model of language understanding described in Chapter 4 is necessarily tailored to the particular constructional formalism of Chapter 3, it nonetheless draws on and adapts many ideas pioneered in early natural language parsing and understanding systems. (See Jurafsky & Martin (2000) for a complete review of the literature.)

Most work in the parsing literature has relied on exclusively syntactic information, typically focusing on context-free grammars. But many early techniques developed to improve efficiency are easily extended to more inclusive domains. These include the use of both top-down and bottom-up cues to direct the search for applicable rules; the ability to look ahead in the input to perform local disambiguation; the use of a *chart* to record previously found units and thereby avoid redundant processing; and the use of *partial parsing* (or chunk parsing) techniques (Abney 1996) to identify islands of certainty within larger sentences not covered by the grammar, increasing robustness to unexpected input. Unification-based parsers have extended such techniques to unification-based grammars, though they often assume a context-free grammatical core whose basic units consist of feature structures. Finally, perhaps the most dramatic innovation in parsing technology has been the rise of probabilistic techniques, especially as applied to lexicalized grammars of various types.

Bryant (2003) describes a unification-based construction analyzer that adapts several of the techniques above, including partial parsing and chart parsing, for use with the constructional domain. In particular, the analyzer performs additional checks to incorporate unification constraints on constructional, form and meaning features. The resulting parse structure (termed an *analysis*) is associated with a set of semantic structures (or *semantic specification*). A version of this analyzer, extended with a form of *reference resolution*, serves as the basis for language understanding in the current model, as described in Chapter 4. In more recent work, Bryant (2008) describes a probabilistic construction analyzer that incrementally finds the best-fitting analysis based on constructional, semantic and statistical cues.

The constructional formalism introduced in the next chapter relies on a unified (paired) representation for form and meaning, based on feature structures. At the constructional level, it thus resembles other unification-based approaches to semantic representation. All of these draw on a longer tradition of *frames*, in the sense of both frame semantics in linguistic theory (Fillmore 1982) (discussed further in Section 3.2.2) and a basic slot-filler structure in knowledge representation (Minsky 1974). The frame-based semantic representations used in the current model are also, however, intended to specify parameters for the structures deployed during simulation, described in the next section.

2.3.3 Embodiment and simulation

A few previous computational models have addressed the embodied nature of linguistic meaning. Several research efforts have focused on the problem of *grounding* language in the physical world by exposing agents (robotic or simulated) to sensorimotor input paired (explicitly or implicitly) with linguistic input. These systems have shown how labels (either speech or text) can become statistically associated with concepts in various semantic domains corresponding to patterns of sensorimotor experience. Some of these use raw audio, video or kinesthetic data as input for language learning object and attribute terms (Roy 1999; Steels 1996; Steels & Kaplan 1998; Steels & Vogt 1997) or verbs (Siskind 2000b; Siskind 2000a; Oates *et al.* 1999). Other models use representations that capture higher-level but nonetheless biologically motivated features in simulated environments. The Regier (1996) model mentioned in Chapter 1, for example, learned spatial prepositions from bitmap representations, with intermediate features similar to those computed by the human visual system.

More recent work in the NTL project has led to the development of a dynamic representation for actions and events appropriate for investigating the Simulation Hypothesis, called an **executing schema**, or **x-schema** (Bailey *et al.* 1997; Narayanan 1997a; Narayanan 1997b). X-schemas are active, graph-based, token-passing structures, formally based on stochastic Petri nets (Reisig 1985) and reducible to structured connectionist models (Shastri *et al.* 1999). They are motivated by features of human motor control, capturing sequential, concurrent and hierarchical events; the consumption and production of resources; and parameterized, context-sensitive execution with variable values. Crucially, x-schemas can be used not merely to *represent* complex actions and events but also to perform (*i.e.*, *execute*) them, either in the physical world (*e.g.*, by a robot) or in a simulated environment. They thus provide a powerful general mechanism for supporting inference through simulation — that is, we can determine the effects and entailments of a given event by actively simulating it and directly inspecting the resulting x-schematic state.

The basic Petri net is a weighted, bipartite graph that consists of *places* (drawn as circles) and *transitions* (drawn as rectangles) connected by directed input and output arcs. The state of a net is defined by a *marking* that specifies a distribution of *tokens* (shown as a black dot or a number) over the places of the net. The real-time execution semantics of Petri nets models the production and consumption of resources: a transition is *enabled* when all its input places are marked such that it can *fire* by moving tokens (the number specified by the weight of the arc) from input to output places. X-schema extensions to the Petri net formalism include typed arcs (modeling resource consumption, enabling conditions or inhibiting conditions); hierarchical control (modeling the decomposition of action hierarchies into subschemas; durative transitions (allowing a delay interval between enabling and firing), parameterization (dynamic binding to tokens representing specific individuals or objects in the environment); and stochastic firing (modeling uncertainty in world evolution or prioritized action selection). The simple x-schema for WALK(TO STORE) shown in Figure 2.2 depicts conditions (such as visual and postural conditions) that allow an agent with sufficient energy to begin an ongoing process of walking by taking a step with each foot, which continues until the agent arrives at the store.

The rich model of event structure afforded by x-schemas has been applied to account for complex phenomena in several linguistic domains, including a wide array of crosslinguistic aspectual distinctions, both lexical and grammaticized, and inferences arising from aspectual composition

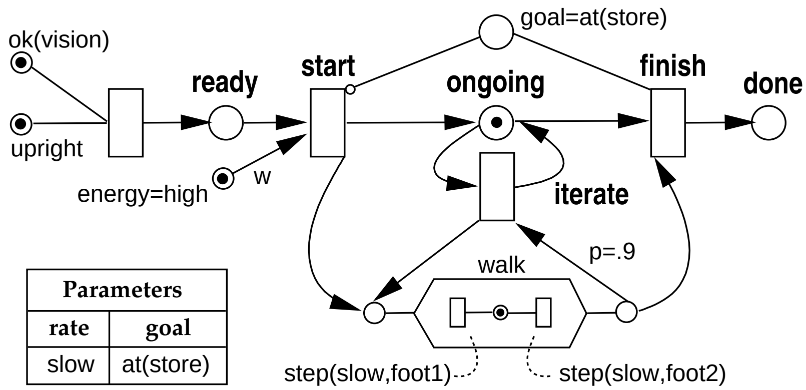

Figure 2.2. X-schema for WALK(TO STORE)

(Narayanan 1997a; Chang *et al.* 1998). Narayanan (1997a) demonstrates how the same model, in combination with a set of metaphorical mappings and dynamic belief nets, supports metaphorical inference based on Lakoff & Johnson's (1980) proposal that metaphorical language is understood in terms of more directly embodied domains like physical motion.

Although the x-schema representation does not play a direct role in the current learning model, it is a crucial computational substrate of simulation-based language understanding. It thus serves as a design constraint on the grammatical formalism that both supports this process and serves as the target learning representation.

2.3.4 Learning as optimization

A variety of previous learning models are consistent with the foundational assumptions of the current model, though mostly for the simplified case of lexical learning. Most lexical learning models, for example, assume at least implicitly that the target of learning is a (constructional) mapping between form and meaning. Many of these employ embodied conceptual representations, such as those mentioned in the previous section. Finally, the idea that linguistic knowledge emerges from the large-scale statistical properties of language use has become something of a guiding precept of computational linguistics, and essentially any data-driven learning model can thus be considered usage-based in this broad sense.

The current task, however, poses more demanding challenges. The domain of multiword constructions requires learning algorithms that can accommodate structured, relational representations. Additional usage-based aspects specific to child language learning include: the *incremental* course of acquisition, with robust performance even at early stages of learning; the importance of individual *instances*, in a sense consistent with exemplar-based learning, and the role of *processes* of language use in shaping the meaning or function associated with individual instances. Some of these issues have been explored in a logical context, for example by Siskind (1997) and subsequent work (Thompson 1998; Thompson & Mooney 1999), which employ relational representations (though for lexical items with strictly compositional semantics); and by Selfridge (1986), which models several aspects of child language and includes rudimentary models of both comprehension and production.

The learning framework adopted here extends the line started by Horning's (1969) probabilistic, optimization-based approach to grammatical induction. This framework seeks to find the optimal grammar given the data, where 'optimal' can be interpreted in a probabilistic context as maximum Bayesian posterior probability, or in an information-theoretic context as *minimum description length* (Rissanen 1989). I defer more technical discussion of these ideas until Chapter 7. Informally, the key idea is that candidate grammars can be evaluated according to a criterion that takes into account both prior expectations about the hypothesis space and observed data. The optimal grammar is thus the one that captures the optimum tradeoff between competing preferences for grammar simplicity and for goodness-of-fit to the data. Crucially, the framework does not specify how candidate grammars are proposed. Thus, the search for grammars might be exhaustive, or constrained by domain-specific heuristics (as in the model to be described).

Optimization-based learning provides a versatile framework that has been applied to various aspects of language learning. Previous research along these lines includes Wolff's (1982) model of the acquisition of syntactic patterns and Goldsmith's (2002) work on crosslinguistic morphological segmentation. The most direct antecedents of the current work, as mentioned in Chapter 1, are based on the *model merging* algorithm. Model merging applies a Bayesian criterion to instance-based learning, where the search strategy involves *merging* (and thereby generalizing) similar examples (Stolcke 1994; Bailey 1997). Stolcke's (1994) model of grammar induction, though neither biologically inspired nor semantically rich, nevertheless addresses the problem of acquiring embedded, multi-unit structures from pairs of sentences and feature-based scene descriptions. Bailey's (1997) model, though limited to single words (or at most rigid two-word templates like *push left*), combines the typological breadth of the Regier spatial relations model with the dynamic semantics of simulation, and provides a plausible connectionist reduction of model merging to recruitment learning (Feldman 1982; Shastri 1988). Both models (described in more detail in Chapter 5), exploit representations that can be used bidirectionally, for both comprehension and production, within their specific domains. Together, they demonstrate the viability of Bayesian model merging for learning a range of probabilistic structures and suggest it may be applied to more linguistically adequate formalisms as well. As we will see, however, the basic model merging strategy must be adapted for current purposes to accommodate the relational representational assumptions of the model and to enforce a tighter connection between language learning and language understanding.

* * *

We have now completed our initial survey of the various disciplinary dialects and customs we may encounter ahead. Despite the internecine struggles engendered by Gold's initial explorations, many points of consensus on the basic problems at hand have been identified, and minority contingents in the linguistic, psychological and computational domains have begun to hold increasingly productive dialogues. The chapters ahead attempt to forge an interdisciplinary coalition in support of the constructional, embodied, usage-based outlook described here, starting in Chapter 3 with a construction-based grammar formalism and the simulation-based model of language understanding it supports.

Chapter 3

Embodied Construction Grammar

What is an idea? It is an image that paints itself in my brain.
— Voltaire

My very photogenic mother died in a freak accident (picnic, lightning)...
— from Lolita, Vladimir Nabokov

Our current enterprise requires a theory of language structure that is compatible with the constraints set forth so far: it must be consistent with the notion of embodiment in language use; it must be formally described and it must be learnable. Embodied Construction Grammar (ECG) is a computational formalism intended to fill this tall order. This chapter provides a general introduction to the formalism, beginning with an overview in Section 3.1 of its role in supporting a broader simulation-based model of language understanding. Section 3.2 and Section 3.3 provide more detailed descriptions of the two basic ECG primitives, *schemas* and *constructions*.

As a relative newcomer in the world of grammatical theories, the ECG formalism remains under active development. Since its original formulation (primarily in Bergen & Chang 2005, with a shorter description in Chang *et al.* 2002a), it has been extended to address various (cross)linguistic, computational and psychological challenges.[1] The version presented here is motivated primarily by the demands of the language acquisition model and thus does not attempt to include all of these

[1] See `http://ecgweb.pbwiki.com` for current developments and publications.

extensions; rather, I focus on the core features needed for representing constituent structure and relational constraints, both within and across the paired domains of form and meaning. This version is sufficient for our current focus on English constructions appearing in early child-directed speech, and it also supports the detailed models of language understanding (Chapter 4) and language learning (Chapter 5) to follow.

3.1 Overview

Embodied Construction Grammar is a grammar formalism designed to support an embodied model of language use. We take as starting point many insights from the construction-based family of approaches outlined in Chapter 2 (Goldberg 1995; Kay & Fillmore 1999; Lakoff 1987; Langacker 1991; Croft 2001). Foremost among these is the observation that linguistic knowledge of all sizes and degrees of specificity, from morphemes to multi-word idioms, can be characterized as **constructions**, or pairings of form and meaning. These form-meaning pairings crosscut traditional levels of linguistic analysis (e.g., phonological, morphological, syntactic, semantic and pragmatic). Along with other construction grammarians, we assume that language users exploit constructions to communicate: during comprehension, to discern from a particular utterance a collection of interrelated conceptual structures; and during production, to generate a surface form that expresses some communicative intent.

ECG is distinguished by its focus on precisely *how* constructions facilitate communication. From this **process-oriented** perspective, it is not sufficient to specify constructional mappings between form and meaning; we must also describe how such mappings interact with world knowledge and the surrounding communicative context to support processes of meaningful language use. These structures and processes are further required to be computationally explicit; that is, they must be described formally enough to support computational implementation.

The current discussion is restricted to language comprehension, although the basic ideas of the approach taken here have analogues in language production. The key functional criterion for successful comprehension is that it allows the hearer to react appropriately to an utterance, whether with language (*e.g.*, by answering a question or responding to a comment), some other kind of action (*e.g.*, by complying with an order or request), or some set of updates to the current belief state. A hearer must therefore glean both the basic propositional content of the utterance and the speaker's intended meaning in context. The difficulties of this task have been amply documented: speakers imply, hedge and underspecify; they obey (and flout) Gricean convention and take poetic license; and even the most prosaic of utterances may be rife with ambiguity and polysemy. While such phenomena fall well beyond the scope of our immediate ambitions, they effectively illustrate how meaning arises from the integration of disparate kinds of information, including constructional mappings, embodied and world knowledge, social and communicative goals, and the dynamically evolving situational context.

The Simulation Hypothesis given in Chapter 1 provides a means of characterizing the interaction between specifically linguistic information (encoded by constructions) and domain-general structures and processes. Figure 3.1 provides an overview of the model:

- Linguistic knowledge consists of a repository of **constructions** that express generalizations linking the domains of **form** (typically, phonological schemas) and **meaning** (conceptual schemas). The meaning domain encompasses a variety of cognitive structures motivated by perceptual, motor, social and other kinds of embodied experience.

- Each **utterance** can be seen as instantiating a set of constructions (and their associated forms and meanings) in a particular **communicative context**. In comprehension, a **constructional analysis** process takes an utterance and its context and determines which constructions the utterance instantiates. The product of analysis is the **semantic specification** (or **semspec**), which specifies which embodied schemas are evoked by the utterance and how they are related. The semspec may undergo a further **contextual resolution** process that links its semantic components to specific contextually available referents.

- The **simulation** process takes as input the (resolved) semspec exploits embodied representations to simulate (or enact) the specified events, actions, objects, relations and states with respect to the current context. The inferences resulting from simulation shape subsequent processing and provide the basis for the language user's response.

The simulation-based architecture has some specific representational consequences, chief among which is that constructions (*i.e.*, "pure" linguistic knowledge) need not bear the entire inferential burden alone, but rather need specify only enough information to run a simulation using sensorimotor and cognitive structures. This division of labor reflects a fundamental distinction between conventionalized, schematic meanings that are directly associated with linguistic constructions, and indirect, open-ended inferences that result from detailed simulation. In effect, constructions provide a limited interface through which the discrete tools of symbolic language can approximate and exploit the multidimensional, continuous world of action and perception.

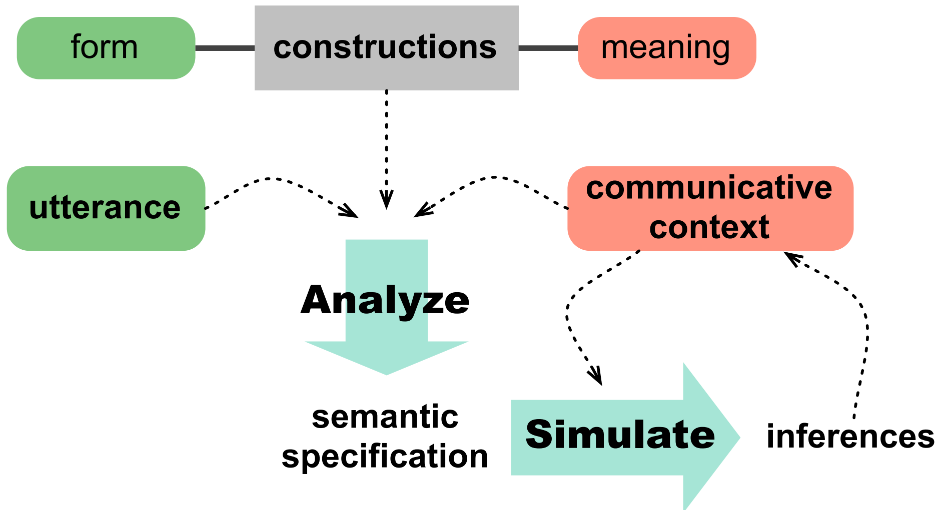

Figure 3.1. Overview of the simulation-based language understanding model: Constructions play a central role as the bridge between phonological and conceptual knowledge, supporting processes of constructional analysis, contextual resolution and embodied simulation.

The remainder of this section introduces the ECG formalism and its use as an interface between surface form and mental simulation, focusing on the representational requirements for understanding a simple example, *Harry ran home*. Section 3.1.1 casts the task in terms of the input sentence and corresponding output simulation, and Section 3.1.2 defines ECG structures that capture basic form, meaning and constructional information relevant to the example. The motivating concern here is to illustrate as simply as possible how the ECG formalism satisfies the broader architectural requirements laid out above. The structures defined are thus intended not to support the most general or linguistically nuanced investigation of the example sentence, but rather to demonstrate how a particular set of linguistic claims might be expressed in ECG. Section 3.1.3 addresses the representational challenges posed by more complex examples; the formal tools available in ECG for meeting these challenges are summarized in Section 3.1.4.

3.1.1 From sentence to simulation

Consider the simple English sentence *Harry ran home.* What surface cues are available to the addressee, and what inferences might be drawn from these? Let us assume that the sentence is pre-segmented into words (regardless of whether it is written or spoken). The surface cues available include, minimally, the individual word forms (*Harry*, *ran* and *home*) and the order in which they appear. We may also include some indication of intonational contour; in its written form the sentence ends with a period, suggesting a neutral or declarative contour for its spoken counterpart.

Based on this limited set of surface cues, a speaker of English is likely to infer that a running event took place sometime prior to speech time; that the runner is an individual named Harry (by default, a human male); and that the goal of running is someone's home (by default, the runner's).

Many other inferences may also be available, though potentially less certain or salient: the runner presumably has legs (at least two), expends energy to move his legs (along with the rest of his body), and starts out at some location other than home. These inferences are not specifically linguistic, but depend rather on general knowledge about runners and running events and homes. Moreover, they are contingent upon the particular referents chosen for *Harry* and *home*: specific interpretive contexts (say, the stipulation that Harry is a baseball player, or that Harry is a horse) might alter the most likely referents of these words, the inferred relationship between them, and the nature of the running motion itself.

Within the framework described here, the bulk of these inferences arise from a mental simulation involving the motor schema for running. Computationally, we represent this simulation using the executing schema formalism described in Section 2.3.3, shown in Figure 3.2 for the relevant running event. The RUN x-schema is similar to the WALK x-schema shown earlier: it is an iterative, energy-consuming process with various parameters, preconditions and effects. Given a particular set of parameters, a simulation based on this x-schema yields a slew of detailed inferences about the event's temporal, causal and interactional structure.

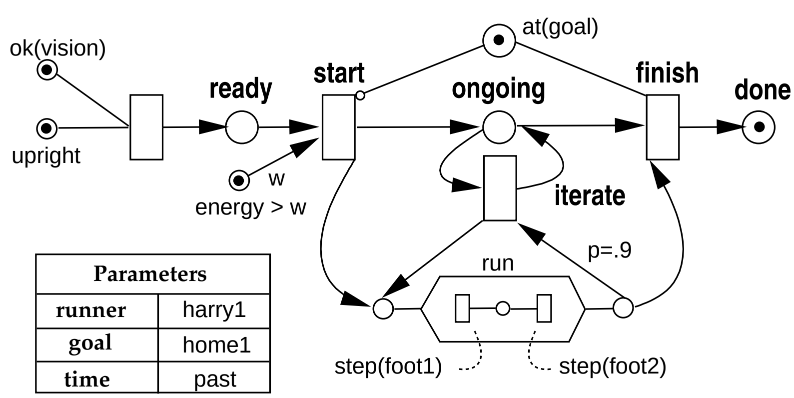

Figure 3.2. An executing schema for running representing a simulation of *Harry ran home*. Tokens in the places corresponding to the conditions done and at(goal) indicate that the goal location has been reached. The parameters harry1 and home1 refer to contextually chosen referents of *Harry* and *home*, respectively, while past indicates that the event takes place before speech time.

We can thus narrow the specifically linguistic portion of language understanding to the following task: given a set of surface cues (corresponding to an utterance) and discourse context (corresponding to its meaning), produce a set of simulation parameters (*i.e.*, the semspec). As noted above, this task divides further into two interrelated processes: constructional analysis, which finds a set of constructions (and associated semspec) that accounts for the input forms; and contextual resolution, which finds a set of contextually appropriate referents for use in the simulation. In this case, the analysis and resolution processes should ideally produce a semspec for a running event with a runner (a referent of *Harry*), a goal location (a referent of *home*), and the time of running (in the past, *i.e.*, prior to speech time); the particular referents resulting from resolution are denoted in Figure 3.2 by the simulation parameters harry1 and home1.

Both of these processes must cope with ambiguity and uncertainty in the input: a given grammar may license many constructional analyses of an utterance, offering competing constituent structures, or varying in degree of constructional specificity; these analyses may in turn afford many contextually licensed resolutions. They thus require some means of evaluating and choosing among candidate analyses and resolutions. Chapter 4 will describe these processes and their associated evaluation strategies in detail. For the purposes of illustrating the ECG formalism, the focus here will be on the representational requirements for supporting a single, simple and unambiguous analysis of the example sentence. The structures defined below thus correspond to an analysis that has minimal internal structure, with no constituent corresponding to the predicate *ran home*; and that is partially lexically specific, treating *home* as a fixed element of the expression, but allowing variation in its other constituents (thus generalizing to sentences like *Theodore sauntered home* and *The dragon is flying home*).[2] More general candidate analyses are considered in Section 3.1.3.

3.1.2 Running example in ECG

Our example analysis, depicted graphically in Figure 3.3, involves three lexical constructions, one for each word in the sentence, along with one phrasal construction, the MOTION-HOME construction. These are shown in the center column of the figure, with arrows from the phrasal construction to the lexical constructions indicating constituency, similar to that in a traditional syntactic parse tree. Each construction also links the form domain (on the left) with the meaning domain (on the right). Here, each lexical construction links a word with a particular conceptual schema, while the phrasal construction links a word order relations (indicated as arrows on the dotted schematic time line) with a set of semantic identity relations (indicated as double-headed arrows).

Such a pictorial representation necessarily elides much detail. Each of the structures shown — including forms, meanings and constructions — is not just an abstract symbol, but rather corre-

[2] This treatment is motivated in part by the idiosyncratic behavior of *home* as a locative noun that can also serve as a path expression (compare with **Harry ran school*).

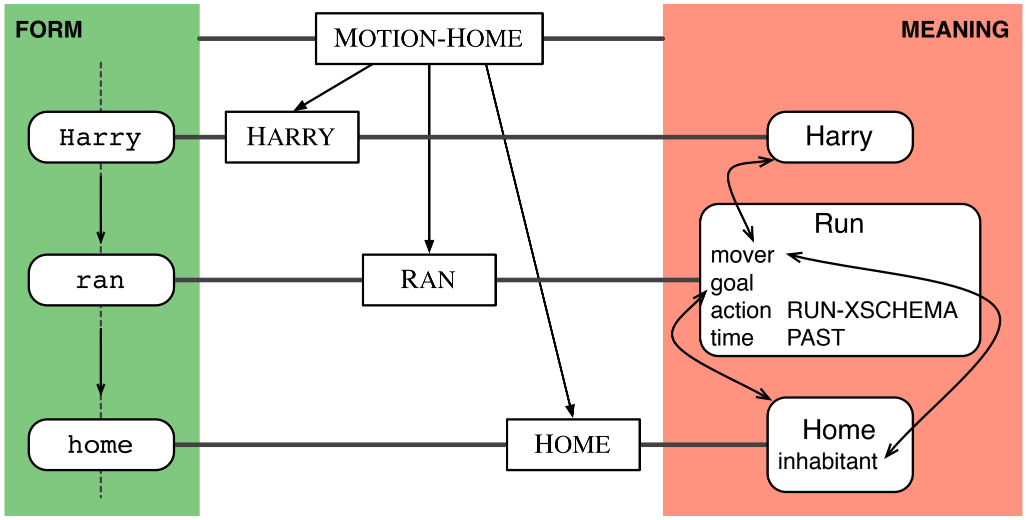

Figure 3.3. A lexically specific analysis of *Harry ran home*. The central boxes depict the constructs involved, each linking schemas in the domains of form (left) and meaning (right). The MOTION-HOME construction asserts constraints on its constituents, including ordering constraints on their forms and identity relations on their meanings.

sponds to a coherent chunk of knowledge within a broader, interconnected (and interdefined) inventory of structures needed to support language use. Below these structures and processes are defined more precisely using the ECG formalism. Our goal is to provide an informal guided tour of the **schema** formalism (for forms and meanings) and the **construction** formalism (for form-meaning maps); see Section 3.1.4 for a detailed summary of the formal tools available in ECG.

The schemas most relevant to our analysis are shown in Figure 3.4. In brief, schemas may have **roles**, corresponding to features or semantic roles (participants, locations, etc.), which may be constrained to be of a specific type (using a colon) or have a particular value (using the $\longleftarrow$ notation). Schemas may also be **subcases** of other schemas, inheriting their roles and constraints.

Summarizing the information expressed by the schemas in Figure 3.4:

- Word has a role phon that is a Phonological-String and a role orth that is an Orthographic-String.
- Harry is a kind of Human whose (inherited) sex role is specified as MALE.
- Home is a kind of Location. Home has a role inhabitant that is Animate.
- SelfMotion has a role mover that is an Entity, a role goal that is a Location, a role path that is an SPG (Source-Path-Goal schema, defined below), a role action that is an X-schema, and a role time that is a Time. Its goal role and mover roles identified, respectively, with the path role's subsidiary goal and trajector.
- Run is a SelfMotion whose (inherited) action role is specified as RUN-XSCHEMA.

The schema formalism is flexible, in that the same notations are deployed to represent, in Figure 3.4(a), general knowledge about words (here, that they are associated with orthographic and

(a)

```
schema Word
  roles
    phon : Phonological-String
    orth : Orthographic-String
```

(b)

```
schema Harry
  subcase of Human
  constraints
    sex ←— MALE
```

```
schema Home
  subcase of Location
  roles
    inhabitant : Animate
```

```
schema SelfMotion
  roles
    mover : Entity
    goal : Location
    path : SPG
    action : Xschema
    time : Time
  constraints
    goal ←→ path.goal
    mover ←→ path.trajector
```

```
schema Run
  subcase of SelfMotion
  constraints
    action ←— RUN-XSCHEMA
```

Figure 3.4. ECG representations of the schemas used in *Harry ran home*, capturing (a) a category for word forms; and (b) conceptual categories contributing to the meaning of the sentence.

phonological forms); and, in Figure 3.4(b), a variety of conceptual structures evoked by the sentence. The formalism is also concise, since many schemas are subcases of others and need only specify any new (non-inherited) roles and constraints. Thus, though the Human schema is not shown, we can assume that its definition is responsible for licensing the inferences that instances of its subcase Harry schema can move, act as a source of energy, have intentions, experience feelings and do whatever else humans do. Likewise, since the Run schema is defined as a subcase of SelfMotion (self-propelled motion), it need only specialize its action role with the appropriate x-schema.

We turn next to the construction definitions needed for the example, beginning with the lexical constructions shown in Figure 3.5. These exploit similar notational devices as the schema definitions, but they include additional structures that reflect their function of mapping form and meaning. The blocks labeled **form** and **meaning** stand for their two linked domains, or **poles** (corresponding to the left and right arrows in Figure 3.3).. These poles list the roles and constraints (if any) within each domain; they should also be considered special roles of the construction, which can themselves be type-constrained. The RAN construction can thus be summarized as follows:

- RAN is a kind of SELFMOTION-VERB.
- RAN has a form pole of type Word with its orth role specified as the orthographic string "ran".
- RAN has a meaning pole of type Run with its time role specified as PAST.

The other two constructions likewise link their form poles (words with specialized orthographic or phonological values) to meaning poles that specialize one of the schemas defined above. Both are also subcases of the REF-EXPR (referring expression) construction, a general construction corresponding to the traditional noun phrase. (Details of this and the SELFMOTION-VERB construction are not important for current purposes; see Section 3.3.)

construction HARRY	construction RAN	construction HOME
subcase of REF-EXPR	**subcase of** SELFMOTION-VERB	**subcase of** REF-EXPR
form : Word	**form** : Word	**form** : Word
orth ⟵ "Harry"	orth ⟵ "ran"	orth ⟵ "home"
phon ⟵ /heri^y/	**meaning** : Run	**meaning** : Home
meaning : Harry	time ⟵ PAST	

Figure 3.5. ECG representations of the lexical constructions used in *Harry ran home*, each linking a specific form with embodied meanings and constraints.

The final construction needed for the analysis, shown in Figure 3.6, combines the lexical constructions into a cohesive gestalt. Like the lexical constructions already described, the MOTION-HOME construction has blocks for describing its form and meaning poles. It also, however, has a block labeled **constituents** for describing its internal constituent structure (as depicted in by the vertical arrows between constructions in Figure 3.3). This block, analogous to the block labeled **roles** in schema definitions, lists its constituent constructions by their local names and associated type constraints. Each of these constituents is itself a construction (or, more precisely, an instance of a construction), with form and meaning poles that are accessible for reference within the larger construction using a subscripted f and m on the relevant constituent name. Moreover, those poles themselves are (instances of) schemas, whose internal roles are also accessible for reference, where dotted notation (x.y) allows reference to accessible structures through a chain of structures. Additional notations are used to impose the form and meaning relations depicted in Figure 3.3: the `before` relation constrains constituent word order; the **identification** operator (indicated by the double-headed arrow ⟷) asserts semantic identity relations; and the keyword **self** allows reference to the structure being defined (here, the MOTION-HOME construction; the subscripted version **self**$_m$ thus refers to the meaning pole of this construction).

The construction definition in Figure 3.6 can be restated as follows:

- MOTION-HOME has a constituent r that is a REF-EXPR, a constituent s that is a SELFMOTION-VERB, and a constituent h that is a HOME.
- The form pole of r comes before the form pole of s, which comes before the form pole of h.
- MOTION-HOME has a meaning pole of type SelfMotion, which is identified with the meaning pole of s. This structure also has a role mover that is identified with the meaning pole of m, and a role goal that is identified with the meaning pole of h.

```
construction MOTION-HOME
  constituents
    r : REF-EXPR
    s : SELFMOTION-VERB
    h : HOME
  form
    r_f before s_f
    s_f before h_f
  meaning : SelfMotion
    self_m ⟷ s_m
    s_m.mover ⟷ r_m
    s_m.goal ⟷ h_m
    h_m.inhabitant ⟷ r_m
```

Figure 3.6. ECG representations of the clausal MOTION-HOME construction, associating word order relations with a self-motion scene.

- The meaning pole of h has a role inhabitant that is identified with the meaning pole of r.

Together, these schema and construction definitions convey the basic information needed to analyze our example sentence. The notational devices provided by the ECG formalism make it possible to express both the relevant domain-specific structures (in form and meaning schemas) and their cross-domain associations (in constructions), making explicit the structures and processes shown in Figure 3.3. Besides elaborating the internal structures of the various schemas and constructions, the formal definitions codify inheritance relations among the structures and allow each structure to specify its locally relevant constraints.

These relations and constraints are combined during the analysis process to produce the constructional graph and its associated set of interrelated schemas. The Run schema illustrates how information from disparate structures is combined: as shown in Figure 3.3, it includes roles inherited from SelfMotion; its action role is specified as RUN-XSCHEMA by the Run schema itself; its time role is specified as PAST by the RAN construction; and its mover and goal roles are bound by the MOTION-HOME construction to the Harry and Home schemas, respectively. This set of bound schemas — *i.e.*, the semspec — supplies the parameters for the requisite simulation using the run x-schema shown in Figure 3.2. The rich set of inferences, encompassing not just the semantic relations expressed by the semspec but also the detailed, context-dependent motor and perceptual consequences of the simulated running event, constitute the meaning of *Harry ran home*.

3.1.3 Beyond *Harry ran home*

We have now illustrated how some simple forms, meanings and constructions are defined using the ECG formalism. These structures support one possible analysis (and corresponding embodied simulation) of our example sentence. This analysis, though pedagogically convenient, is by no means the only one possible, nor necessarily the best. Alternative analyses might involve purely cosmetic variations (perhaps with a more perspicuous choice of names for the relevant schemas, roles and constituents); or they might require more substantive changes to constructional level of specificity or to inheritance relations among structures. For example, we could replace the MOTION-HOME construction with the even more lexically specific HARRY-RAN-HOME construction (with constituents typed as instances of HARRY, RAN and HOME). Such a construction, while of more limited utility than the MOTION-HOME construction defined earlier, would result in the same set of schemas and bindings; *i.e.*, it would be *informationally equivalent* with respect to simulation.

The space of possible analyses grows more daunting as we look beyond our introductory example. Consider the minimal variants shown in (3–1):

(3–1) a. {The boy / He / My oldest son} ran home.
b. Harry {is running / will walk / has flown} home.
c. Harry ran {out / here / *house / to the store / around the house}.

These sentences could be analyzed as having the same high-level constituent structure as our example. In fact, given some additional constructions to license the wider range of expressions for the relevant constituents, the variants in (3–1a) (referring expressions) and (3–1b) (tense and aspect constructions) could make use of the MOTION-HOME construction. The variants in (3–1c), however, necessitate a more general treatment of path expressions, with subcases corresponding to directional particles (like *out*), spatial prepositional phrases (like *to the store* or *around the house*), and locative expressions that express a path (like the locative pronoun *here*).[3] (A number of the requisite constructions are defined in Section 3.3.)

Different communicative functions may also be expressed with respect to our main event, as exemplified in (3–2). These include questions about the veracity of the event as a whole or its individual role-filler values, as well as commands and focus-manipulating constructions:

(3–2) a. Did Harry run home?
b. Who ran home?
c. Where did Harry run?
d. Run home, Harry!
e. It was Harry who ran home.

Venturing past the confines of Harry's running activities, we observe instances of some other basic clausal types in English, corresponding to the main argument structure constructions discussed by Goldberg (1995):

[3]The idiosyncratic behavior of *home* (compared with other common nouns) may justify its inclusion in this last category.

(3–3) a. Mary threw a ball. *(transitive scene)*
b. Mary threw a ball to me. *(caused motion scene)*
c. Mary threw me a ball. *(transfer scene)*

All of these variations raise issues of linguistic representation deliberately sidestepped in our initial example. In particular, while many potential grammars might be informationally equivalent with respect to these examples (in the sense mentioned above of producing the same semspec), they may vary along many other dimensions. A standard criterion employed in linguistic analysis favors grammars that capture shared structure and (relatedly) generalize to further examples; for example, a general SUBJECT-PREDICATE construction is typically assumed to capture commonalities across the sentences in (3–3), such as possible constraints on agreement and word order. In some circumstances, however, one might prefer a grammar that involves fewer levels of analysis (*e.g.*, one in which the examples in (3–3) are analyzed using three wholly independent clausal constructions).

The ECG formalism is, by design, expressive enough to support a wide range of linguistic claims and analyses, but agnostic about which of these should be considered optimal. Indeed, as noted earlier, nothing in the formalism precludes multiple competing or compatible analyses. In principle, the best analysis in a given situation may depend on a host of factors, including dynamic aspects of the analysis and simulation context; the applicability of crosslinguistic, psychological, aesthetic or practical considerations; and, most relevantly for the current work, criteria based on processes of language learning. ECG's theoretical commitments thus apply mainly to the structural level — that is, to the *kind* of information that must be represented to capture the interdependence of linguistic structure, process and context inherent in our approach to language understanding.

3.1.4 The formalism at a glance

This section briefly summarizes the notational devices available in the ECG formalism; many of these have been introduced informally in the preceding sections, and all are illustrated with more examples in Section 3.2 and Section 3.3. See Chapter 4 for more formal definitions.

Computationally, the formalism draws most directly from constraint-based grammars of the kind discussed in Section 2.3.1 (Shieber 1986; Pollard & Sag 1994): the underlying representation is a typed feature structure with unification constraints and feature inheritance. But, as illustrated in the last section, the formalism bears a closer surface resemblance to object-oriented programming languages like Java, with special keywords and symbols for expressing relations and constraints among structures and allowing self-reference and reference to non-local structures.

The version of the formalism presented here simplifies many aspects of linguistic representation. As discussed in Section 2.2.1, constructions are like other complex radial categories in exhibiting graded membership and internal structure; the multiple inheritance hierarchy assumed here captures some but not all aspects of constructional category structure. In particular, graded category membership might be better expressed in a probabilistic framework (as discussed again in Chapter 9). The simpler inheritance-based organization used here, however, suffices for our current focus on structural relations among constructions.

Representational primitives

ECG includes formalisms for several representational **primitives** needed for capturing a range of linguistic phenomena; only the two most relevant of these are discussed here:

- A **schema** is the basic structured unit of representation for both the form and meaning domains. Each schema specifies relationships among a set of interdefined participant **roles**. Roles can be instantiated by particular values (or *fillers*). Form schemas provide information relevant to surface form (*e.g.*, associated phonological or orthographic strings, intonational information, temporal ordering), while meaning schemas help to specify parameters for embodied simulations.

- A **construction** is the basic linguistic unit that pairs elements and constraints in the form and meaning domains, or **poles**. Each construction has a **form** pole and a **meaning** pole, which can be constrained to instantiate specific form and meaning schemas, respectively. Some constructions also have internal **constituents** that are themselves constructions. Constructions may also have features encoding properties of the construction that do not reside solely in the form or meaning domain.

The other primitives are *spaces* and *maps*; see (Chang *et al.* 2002a; Mok *et al.* 2004) for discussion of how these are used to account for metaphorical mappings and mental space phenomena.

Relations among primitives

Structures may be related in several ways:

- **subcase**: Schemas and constructions are organized in multiple inheritance hierarchies, each a partial order induced by the **subcase** relation between a structure (schema or construction) and its more general **base** structure (or set of base structures), notated using the tag **subcase of** x. The subcase relation is similar to the *is-a* link used in semantic networks. The roles (and constituents, in the case of constructions) of each base structure are accessible to its subcases, and its constraints apply.

- **constituency**: Complex schemas and constructions may combine a set of simpler subsidiary structures. A schema may contain *roles* instantiating other schemas, and a construction may contain *constituents* instantiating other constructions. This relationship corresponds to a *part-whole* relation, allowing one schema or construction to include another.

- **evocation**: A schema may *evoke* (or activate) an instance of another schema x with the local identifier y, without implying either inheritance or constituency, using the notation **evokes** x

as y. This underspecification provides needed flexibility for building semantic specifications. In combination with the **self** notation (see below), it also allows one structure to be defined or raised to prominence against a background set of structures, formalizing the notion of *profiling* used in frame semantics (Fillmore 1982) and Cognitive Grammar (Langacker 1991).

- **containment**: In addition to constituency, a construction may be seen as containing its form and meaning poles, as well as any additional roles it introduces in the form, meaning or constructional domains.

Accessible structures

ECG includes notations for expressing several kinds of constraints. Arguments to these constraints must be *accessible* structures within the relevant definition, *i.e.*, one of the following:

- the structure itself, expressed using the keyword **self**;
- locally defined roles, constituents and evoked structures;
- inherited roles, constituents and evoked structures (*i.e.*, any structures accessible via the subcase relation);
- roles and constituents recursively accessible through other accessible structures, using a dotted "slot chain" notation to refer to a role y of a structure x as $x.y$; and
- the form and meaning poles of any accessible construction, including those of the structure itself or any of its constituents, using a subscripted f or m.

Constraints

The following constraint types are allowed:

Type (or **category**) constraints (indicated with a colon, as $x : y$) restrict x to be filled by an instance of schema y.

Binding constraints: ECG has two constraints that correspond to standard unification or coindexation. **Identification** constraints (indicated with a double-headed arrow, as $x \longleftrightarrow y$) cause fillers to be shared between x and y, thus indicating how roles of different structures involved in simulation are aligned. **Filler** constraints (indicated with a single-headed arrow, as $x \longleftarrow y$) indicate that the role x is filled by the element y (a constant value).

Ordering constraints: Temporal relations among form segments are notated using form constraints. In principle, the formalism can express any binary relation between intervals, including sequence, overlap, contiguity and others specified in Allen's (1984) Interval Algebra. Of these, the most common relations are **before** (for precedence) and **meets** (for immediate precedence). In the absence of any explicit order constraint, a weaker co-occurrence constraint holds among the forms of different constituents of the same construction.

Two other constraint types are allowed, as described elsewhere (Chang *et al.* 2002a; Chang *et al.* 2002b): *predicates*, a general means of expressing open-ended relational constraints; and *dynamic constraints* that allow other constraints to be asserted as holding only within a specific setting (temporal, spatial, etc.). These are not exploited in the current work.

*   *   *

The next two sections elaborate on the formalism's expressive capacity. For expository ease (and maximal relevance to the learning model), these focus on the representational demands posed by crosslinguistic patterns of lexicalization and grammaticization common in early child language.

3.2 Schemas

Embodied schemas generalize over a long tradition of representations in linguistic analysis. These include feature-based representations for phonological, morphological and syntactic variation, as well as the role-filler semantic relationships associated with case-marking and thematic roles. All of these are schematic structures based on structured relations over variable values. ECG schemas are distinguished from other feature-based representations in two key respects: (1) ECG schemas are explicitly associated with richer underlying structures: schema roles are viewed as supplying both parameters to the associated grounded representations and points of access for other schemas and constructions. Thus, they can function not only as symbols that name relationships (as features often do) but, more importantly, as an interface between structures at different levels of granularity and across different domains. (2) ECG schemas include additional notational devices that make it convenient to express both traditional role-filler bindings and the more complex relationships that arise within interdefined clusters of concepts.

The sections below describe the main varieties of embodied schemas relevant to the current work. These include schemas from the basic domains of form (Section 3.2.1) and meaning (Section 3.2.2), as well as a special class of schemas for capturing communicative context (Section 3.2.3).

3.2.1 Form

The canonical domain of linguistic form is sound. The crosslinguistically most common sound patterns are described abstractly in terms of phonological sequences, tone and stress, or more concretely as the acoustic patterns and the articulatory gestures that produce them; written languages also employ orthographic representations. Form in its most general sense, however, may extend to any kind of signifier, including manual and facial gestures, pictorial icons like those taught to non-human primates, and tactile patterns like those in Braille.

Relationships among forms vary by domain. Many of these include a temporal dimension (or a spatial dimension with a temporal correlate), and thus naturally afford certain relations among

```
schema Schematic-Form
```

```
schema Word
  subcase of Schematic-Form
  roles
    phon : Phonological-String
    orth : Orthographic-String
```

```
schema Complex-Word
  subcase of Word
  roles
    stem
    prefix
    suffix
  constraints
    prefix before stem
    stem before suffix
```

Figure 3.7. Form schemas: The Schematic-Form schema is the most general form schema; its (simplified) subcase Word schema has roles for phonological and orthographic representations.

temporal segments or intervals: *e.g.*, one segment may precede one another, either immediately or with intervening segments; segments may include other segments, possibly as initial or final segments; two segments may be temporally coincident (Allen 1984). Such interval relations are sufficient to describe many of those associated with concatenative morphology (namely, affixation) and syntax (word order). Other phenomena, like templatic morphology, vowel shifting and vowel harmony, may require domain-specific similarity- and variation-based relations, or even the coordination of multiple forms in different domains (such as the linked manual and facial gestures in some sign languages).

The ECG schema formalism can accommodate any of the forms and relations above. The current work makes use of a small subset of the possible forms and relations, focusing on orthographic information, as in the introductory example. Figure 3.7 shows some example form schemas, including the abstract Schematic-Form schema of which all other form schemas are subcases; a Word schema (repeated from Section 3.1.1) with roles for phonological and orthographic strings; and a Complex-Word schema that might be useful for languages with concatenative morphology.

3.2.2 Meaning

We take the scope of linguistic meaning to include all conceptualizable notions, from concrete entity, relation and action categories to the more abstract categories usually associated with grammar. It is thus appropriate that embodied schemas for meaning generalize over many proposed semantic representations. The basic notion — that of a schematic structure capturing relations among a set of roles that can be filled by variable values — is similar to the *semantic schema* assumed in Langacker's (1987) Cognitive Grammar and compatible with Lakoff's (1987) *idealized cognitive model* and Fillmore's (1982) *frame* (as discussed further below).

Figure 3.8 shows schemas representing some standard examples from the literature. As depicted here, both concrete physical categories like those in (a) and specific named individuals like those in (b) are represented as schemas, but the latter are also defined as subcases of a NamedEntity category. The action schemas in (c) specialize the SelfMotion schema defined earlier, and (d) includes some commonly grammaticized image schemas, such as Trajector-Landmark schema (spatial or attentional asymmetry), Container (enclosed or partially enclosed regions) and SPG (motion along a path) (Johnson 1987; Lakoff & Johnson 1980).[4]

(a)

```
schema Human
  subcase of Animate
  roles
    sex
    status
```

```
schema Girl
  subcase of Human
  constraints
    sex ⟵ FEMALE
    status ⟵ CHILD
```

```
schema Cafe
  subcase of Location, Business
  constraints
    product ⟵ COFFEE
```

(b)

```
schema Harriet
  subcase of Girl, NamedEntity
```

```
schema Starbucks
  subcase of Cafe, NamedEntity
```

(c)

```
schema Walk
  subcase of SelfMotion
  roles
    mover : Animate
  constraints
    action ⟵ WALK-XSCHEMA
```

```
schema Hop
  subcase of SelfMotion
  roles
    mover : Animate
  constraints
    action ⟵ HOP-XSCHEMA
```

(d)

```
schema Trajector-Landmark
  roles
    trajector
    landmark
    relation
```

```
schema Container
  roles
    interior
    exterior
    portal
    boundary
```

```
schema SPG
  roles
    source : Location
    path : Path
    goal : Location
    trajector : Entity
```

Figure 3.8. Meaning schemas: (a) entities; (b) named entities; (c) actions; (d) image schemas.

A key feature of semantic representation, associated most closely with the study of *frame semantics* (Fillmore 1982), is that many concepts are defined against a background set of related participants and props (or *frame*). The classic example of a culturally specific frame is that of the commercial transaction, *i.e.*, an exchange of goods and money between a buyer and seller. Different aspects of this inferentially rich backdrop are highlighted, or *profiled*, by lexical items like *buy*, *sell*, *price*, *cost* and *discount*. Langacker (1987) provides a geometric example exhibiting the same representational pattern: the concept associated with the word *hypotenuse* cannot be defined independently of

[4] As noted in 2.2.2, a variety of terms appear in the literature for related concepts. Alternative terminology more in line with Talmy's (2000) schematization might include the Figure-Ground schema (with roles figure and ground) instead of Trajector-Landmark and a Path schema (with roles figure, ground, path and manner) instead of SPG. Our purpose here is not to choose among schematizations but to illustrate that the formalism can accommodate either.

its base concept of a right triangle. The ECG schema formalism expresses this profile-base relation using the **evokes** notation, which allows a schema to evoke a background structure (corresponding to the background frame or base schema). Roles of this evoked structure can then be bound as appropriate; in particular, they may be bound to the current structure (denoted by the **self** notation). Schemas for Hypotenuse and its evoked schema Right-Triangle are shown in Figure 3.9. Note that this approach to profiling can also be used within the meaning pole of a construction definition, as illustrated earlier by the HOME construction in Figure 3.5.

```
schema Right-Triangle
 roles
   leg1 : LineSegment
   leg2 : LineSegment
   hyp : LineSegment
   angle1 : Angle
   angle2 : Angle
   angle3 : Angle
 constraints
   angle3 ⟵ 90
```

```
schema Hypotenuse
 evokes Right-Triangle as rt
 constraints
   self ⟵ rt.hyp
```

Figure 3.9. Profiling by evoking in ECG: The Hypotenuse schema **evokes** the Right-Triangle and binds itself to (or profiles) one of its roles.

Complex interactional patterns combine the notational devices illustrated above. For example, the word *into* involves a dynamic spatial relation in which one entity moves from the exterior to the interior of another. This situation is depicted informally in Figure 3.10; the corresponding ECG schema is shown in Figure 3.11. In brief, Into is defined as a particular kind of trajector-landmark relationship, one whose landmark is an instance of the Container schema and whose trajector is also the trajector of an evoked instance of the SPG schema. Identification constraints assert that the trajector's path takes it from the exterior to the interior of the container. (The same evoked schemas with a different set of bindings would be needed to express the meaning of *out of*.)

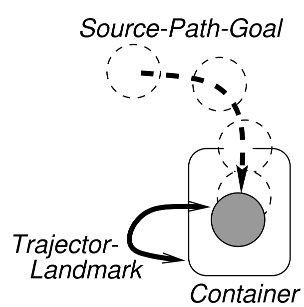

Figure 3.10. An iconic representation of the meaning of *into*, combining the simpler schemas Container, Trajector-Landmark and Source-Path-Goal.

```
schema Into
  subcase of Trajector-Landmark
  evokes SPG as spg
  roles
    trajector : Entity
    landmark : Container
  constraints
    trajector ⟷ spg.trajector
    spg.source ⟷ landmark.exterior
    goal ⟷ landmark.interior
```

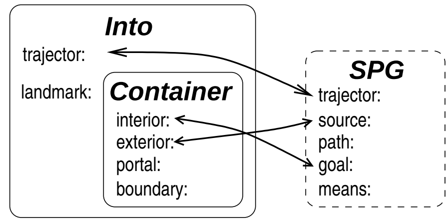

Figure 3.11. The Into schema, defined using the ECG formalism (left) and informally depicted as a linked set of schemas (right). The Into schema is defined here as a subcase of Trajector-Landmark that evokes an instance of the SPG schema (shown with a dashed boundary at right).

3.2.3 Function

Utterances are rooted in specific situational and discourse contexts and fulfill specific discourse functions: speakers pick out individuals, indicate whether they have been previously mentioned, make assertions and ask questions about them, and so forth (as demonstrated by the examples in (3–2) from Section 3.1.3). The ECG formalism defines several special schemas that allow explicit mention of the current context and invoke the basic communicative functions of *reference* and *predication*, as shown in Figure 3.12:

- The current discourse context, or **discourse space**, can be characterized in terms of a set of interlocutors and other salient participants, and the surrounding attentional and activity context (as in the DiscourseSpace schema).

- **Reference** is a basic communicative function in which the speaker evokes or directs attention to specific entities and events (the **referent** or set of referents). Languages differ in precisely what kinds of information speakers supply; as shown in the Referent schema, these typically include the referent's ontological category (*e.g.*, human, ball, picture), number and gender; its default level of *accessibility* (Lambrecht 1994) in the current context (active, accessible, inactive, unidentifiable, etc.); and any restrictions and attributions that apply to its open-class characteristics (such as size or color). The last role shown, the resolved-referent, is a special role filled by the actual resolved entity described by the schema.

- **Predication** is the relational counterpart to reference. Speakers make attributions and assert relations as holding of particular entities; and they locate, or ground, these relations (in time and space) with respect to the current speech context. The Predication schema provides roles for an overall scene (typically set by an argument structure construction) and its central means (typically set by verbal constructions); the predicated event's protagonist (main participant), event-structure (aspectual profile) and temporal or spatial setting; and any associated speech act type.

```
schema DiscourseSpace
  roles
    speaker
    addressee
    joint-attention
    activity
```

```
schema Referent
  roles
    category
    number
    gender
    accessibility
    restrictions
    attributions
    resolved-referent
```

```
schema Predication
  roles
    scene
    means
    event-structure
    time
    place
    speech-act
```

Figure 3.12. Schemas for describing communicative function

These structures play a central structuring role in grounding schemas with respect to the current context (by constraining the contextual resolution process described further in Chapter 4) and setting parameters for the simulation to be executed. But, as shown here, they can be schematized and represented using the ECG schema formalism, thus allowing them to mingle freely with other constructional meanings.

* * *

The examples above demonstrate the expressive capacity of the ECG schema formalism. Again, we make no strong claims about the particular schemas defined here, which might be modified or extended, for example, to support crosslinguistic variation, or reflect other task-specific assumptions. The formal tools they illustrate are, however, adequate for describing a wide range of form, meaning and function schemas, including those most relevant in early language learning.

3.3 Constructions

Constructions serve as the crucial symbolic link between the domains of form and meaning. Each construction is associated with two *domain poles*, representing these two domains. These poles are formally similar to ECG schemas: they may be typed as instances of existing schema types, and they may be further specialized with additional constraints. Constructions also resemble schemas in that they are defined within a multiple inheritance hierarchy, and they may have additional constituent structure, analogous to subsidiary schema roles. The ECG construction formalism thus employs many of the notations used for schemas, along with additional mechanisms for describing cross-domain associations. This section surveys the varieties of cross-domain mapping permitted in linguistic constructions (Section 3.3.1) and summarizes some basic construction types (Section 3.3.2 and Section 3.3.3).

3.3.1 Cross-domain mapping types

As previously described, constructions cut across traditional linguistic divisions: they vary in both size (from morphological inflections to intonational contours) and level of concreteness (from lexical items and idiomatic expressions to clausal units and argument structure patterns). Formally, however, the most salient difference among constructions is based on the presence or absence of internal constituent structure. A **simple** construction consists of only one cross-domain mapping. That is, though both the form and meaning domains may contain arbitrarily complex sets of schemas, the mapping between them itself lacks internal structure. In contrast, a **structured** (or **complex**) mapping includes at least one subsidiary cross-domain mapping, corresponding to a constituent.

Simple and structured mappings correspond canonically to lexical and phrasal (or clausal) constructions, respectively. Our introductory example in Section 3.1.1 included three simple, lexical constructions (for *Harry*, *ran* and *home*) and one structured mapping combining them. Figure 3.13 illustrates a similar pattern using constructions that license the phrase *into Starbucks*. The two lexical constructions in (a) correspond to the preposition *into* and the proper noun *Starbucks*. Each of these links a single form schema (a word with an orthographic constraint) with a single meaning schema (taken from Figure 3.8). In (b) we show the schematic PATH-EXPR (pairing a form of unspecified type with a path description) and its subcase PATH-PHRASE construction, an ECG version of the familiar prepositional phrase. Separate constituents denote the particular spatial relation involved (PATH-PREP) and a landmark referent; the form poles of these constituents are constrained to come in a particular order, and their meanings are aligned as indicated. Specifically, the Trajector-Landmark relationship denoted by the prepositional constituent rel has a landmark role (rel_m.landmark) that is identified with the referent of the other constituent (lm_m). The construction's overall meaning is also bound to the (evoked) spg role of the prepositional constituent (rel_m.spg), an instance of the SPG schema (as defined in Figure 3.11).

The distinction between simple and structured mappings applies to constructions of all levels

(a)

construction INTO-CXN
subcase of PATH-PREP
form : Word
orth $\longleftarrow$ “into”
meaning : Into

construction STARBUCKS
subcase of REF-EXPR
form : Word
orth $\longleftarrow$ “Starbucks”
meaning : Starbucks

(b)

construction PATH-EXPR
form : Schematic-Form
meaning : SPG

construction PATH-PHRASE
subcase of PATH-EXPR
constituents
rel : PATH-PREP
lm : REF-EXPR
form : Phrase
rel_f `before` lm_f
meaning : SPG
rel_m.landmark $\longleftrightarrow$ lm_m
$\textbf{self}_m$ $\longleftrightarrow$ rel_m.spg

Figure 3.13. Simple and structured constructional mappings: (a) simple lexical constructions for INTO and STARBUCKS; (b) the PATH-PHRASE construction, a subcase of PATH-EXPR with constituents capturing a structured mapping.

of abstraction and size. Lexical items may have internal (morphological) constituents, and multi-word phrases may be linked *in toto* by a simple mapping to an associated (and perhaps frozen idiomatic) meaning. In fact, both kinds of mappings for the same construction can co-exist: the preposition *into*, defined as a simple map in Figure 3.13, could also be described as a compound with separate semantic contributions from *in* and *to*; the phrase *out of* offers a similarly rich set of possible analyses. Multiple mappings may also be useful for capturing partial idiomaticity, like that in the idiom *kick the bucket*: the phrase as a whole maps non-compositionally to a simple meaning like that of the word *die*, but a separate simple mapping for the verb *kick* could license variation based on tense-aspect-modality constructions. As we will see in Chapter 6, this ability to represent multiple kinds of mappings allows initially simple representations to be reanalyzed and enriched over the course of development.

3.3.2 Reference and predication

So far we have shown scattered examples of constructions needed for simple examples. We now embark on a more organized tour of some basic English constructions. Though neither exhaustive nor definitive, this overview exemplifies the main relationships among form, meaning and function expressible by the ECG construction formalism. We analyze many traditional grammatical categories and relations as motivated by the primary propositional acts of reference and predication

(Croft 1990; Croft 1991; Croft 2001).[5] That is, they are defined in terms of their communicative functions, as represented by the Referent and Predication schemas defined in Section 3.2.3. Figure 3.14 shows the schematic REF-EXPR and PRED-EXPR constructions, which generalize over all referring and predicating expressions, respectively. Some constructions are defined as subcases of these, linked directly to the relevant communicative schemas; others use them as background frames against which specific roles are profiled. We illustrate some of the major construction types below.

construction REF-EXPR	construction PRED-EXPR
form : Schematic-Form	**form** : Schematic-Form
meaning : Referent	**meaning** : Predication

Figure 3.14. Schematic constructions for referring and predicating expressions

Referring expressions

A wide range of expressions draw on the notion of reference. Some of these function as full referring expressions — *i.e.*, they map a simple or complex form directly to the Referent schema. Figure 3.15 shows two referring expressions, both simple subcases of REF-EXPR:

Proper nouns pick out specific named entities (by default) and can typically be used with no prior mention. In the STARBUCKS construction (which elaborates the version in Figure 3.13), the resolved-referent role is directly bound to the specified schema, and only a minimal or inactive level of accessibility is required (*i.e.*, no prior mention is generally needed).

Pronouns pick out contextually available referents (*i.e.*, ones for which the interlocutors have active representations in the current discourse) and assert additional constraints on features like number and gender, as in the HE construction. In this example, we include a case constructional feature (in a **constructional** block) to distinguish the HE construction from the constructions for *him* (object case) and *his* (possessive case). Languages with wider use of case may define this feature for many other constructions.

More complex referring expressions require structured mappings, as well as auxiliary constructions that constrain specific aspects of the referent. Figure 3.16 illustrates constructions that license the phrase *the cafe*:

[5] Croft identifies *modification* as a third category, exemplified by adjectival and adverbial constructions. We analyze these instead as evoking the same structures as involved in reference and predication and asserting further constraints, but a general constructional category for such expressions could also be defined.

construction STARBUCKS (2)
subcase of PROPER-NOUN
form : Word
orth ⟵ "Starbucks"
meaning
accessibility ⟵ inactive
resolved-referent ⟷ Starbucks

construction HE
subcase of PRONOUN
constructional
case ⟵ subject
form : Word
orth ⟵ "he"
meaning
accessibility ⟵ active
number ⟵ singular
gender ⟵ male

Figure 3.15. Referring expression examples: The proper noun STARBUCKS and pronoun HE constructions bind the Referent schema's resolved-referent role and set default levels of accessibility. The HE construction also constrains its case constructional feature.

construction DETERMINER
form : Schematic-Form
meaning
evokes Referent **as** ref

construction COMMON-NOUN
form : Schematic-Form
meaning
evokes Referent **as** ref
self$_m$ ⟷ ref.category

construction THE-CXN
subcase of DETERMINER
form : Word
orth ⟵ "the"
meaning
ref.accessibility ⟵ accessible

construction CAFE-CXN
subcase of COMMON-NOUN
form : Word
orth ⟵ cafe
meaning : Cafe

construction DETERMINEDNOUN-EXPR
subcase of REF-EXPR
constructional
det : DETERMINER
com-noun : COMMON-NOUN
form
det$_f$ `before` com-noun$_f$
meaning
self$_m$ ⟷ com-noun$_m$.ref
self$_m$ ⟷ det$_m$.ref

Figure 3.16. Determined noun construction: Determiners and common nouns evoke and constrain a Referent schema; they are combined in the DETERMINEDNOUN-EXPR construction.

Determiners assert various constraints on a referent, for example on its number or contextual accessibility. The construction for *the* constrains the latter of these to be accessible; analogous constructions for *a* or *some* would constrain the referent to be inaccessible and singular or plural, respectively.

Common nouns denote categories; the COMMON-NOUN construction evokes a Referent and binds its meaning pole to the category role. Subcases like *cafe* or *boy* specialize their form and meaning poles with appropriate phonological or orthographic strings and semantic categories.

Determined nouns combine a determiner with a category-denoting expression (such as common nouns) to pick out an individual instance of the specified category. A simple case of this is a phrase like *the cafe*, referring to a specific member of the conceptual category of Cafe that has been previously mentioned or whose identity is otherwise inferrable.

This treatment of reference, though preliminary, has been successfully extended to modifiers, complements, quantifiers and measure phrases. While further research is necessary, we believe the basic strategy of evoking and constraining the Referent schema can scale to accommodate a wide full range of referential phenomena.

Predicating expressions

Constructions supporting the basic function of predication are represented using a strategy similar to that employed for reference: the Predication schema serves as the background frame whose roles specify what kind of action or relation is being asserted, commanded or queried. Various kinds of constructions can contribute to an overall predication by evoking this schema and binding it as appropriate; the unification of these evoked predications forces them to be understood as — *i.e.*, construed as — part of a single coherent event.

Verbs and verb complexes designate relations among entities. We define verbs as evoking a predication and binding to (or profiling) its central schema (or means). Subcases might specialize the action or event category (as do the SelfMotion-Verb and Walk constructions in Figure 3.17) and assert additional constraints on tense (locating the predication with respect to the current speech time, as in the Walked construction), aspect (event structure) or modality (relation to reality).

```
construction Verb
  form : Word
  meaning
    evokes Predication as pred
    self_m ⟷ pred.means
```

```
construction SelfMotion-Verb
  subcase of Verb
  form : Word
  meaning : SelfMotion
```

```
construction Walk
  subcase of SelfMotion-Verb
  form
    orth ⟵ "walk"
  meaning : Walk
```

```
construction Walked
  subcase of Walk
  form
    orth ⟵ "walked"
  meaning
    pred.time ⟵ past
```

Figure 3.17. Basic verb constructions: The Verb construction and its subcases SelfMotion-Verb, Walk and Walked.

Argument structure constructions specify a basic pattern of interaction among a set of participants within a *scene* (Goldberg 1995). The ArgStruct construction shown in Figure 3.18 does this by evoking a predication and profiling its scene role; this predication must also unify with that evoked by its required verbal constituent. Subcases of ArgStruct may declare additional constituents, specialize existing constituents, and assert further constraints. The SelfMotion-Cxn, for example, adds a constituent corresponding to the mover entity, specializes the verbal constituent to require a Motion-Verb, and asserts constraints on case, word order and role bindings. Other argument structure constructions corresponding to the basic scenes discussed in Section 2.2.1 (*e.g.*, caused motion, transitive, transfer) may likewise add, specialize and constrain constituents as appropriate.

```
construction ArgStruct
  constituents
    v : Verb
  form : Phrase
  meaning
    evokes Predication as pred
    self_m ⟷ pred.scene
    v_m.pred ⟷ pred
```

```
construction SelfMotion-Cxn
  subcase of ArgStruct
  constituents
    a : Ref-Expr
    v : Motion-Verb
    a.case ⟵ subject
  form
    a_f before v_f
  meaning : SelfMotion
    mover ⟷ a_m
    action ⟷ v_m
```

Figure 3.18. Basic argument structure constructions: The ArgStruct construction and its subcase SelfMotion-Cxn.

This treatment allows a more general analysis of *Harry ran home*, along the lines discussed in Section 3.1.3. The DirectedMotion-Cxn construction in Figure 3.19 is defined as a subcase of SelfMotion-Cxn with an additional argument expressing the path.[6] Unlike the lexically specific Motion-Home construction in Figure 3.5, the DirectionMotion-Cxn allows a variety of path expressions to specify the motion's direction. Note, however, that both kinds of constructions can coexist in a grammar, and neither is intrinsically better than the other: the more general construction accounts for more variations, but the more specific construction is more predictive of those variants it licenses. It is precisely this tradeoff that drives the construction learning model to be described.

The examples given here do not do justice to the varieties of predicating constructions posited in the literature, such as those that capture generalizations about subject-predicate agreement, active-passive voice and question formation. These fall beyond the scope of our current focus on child language learning.

[6] The analysis in Section 3.1.1 is also simplified in that the relevant constraints contributed by the verb (action and time) and clausal construction (scene) are asserted directly on the SelfMotion scene, in lieu of an evoked Predication.

```
construction DirectedMotion-Cxn
  subcase of SelfMotion-Cxn
  constituents
    b : Path-Expr
  form
    v_f before b_f
  meaning
    path ⟷ b_m
```

Figure 3.19. The DirectedMotion-Cxn, a subcase of SelfMotion-Cxn

3.3.3 Other constructions

The constructions shown so far cover only basic aspects of reference and predication. The principles of linguistic representation they exemplify can, however, be extended to other constructional domains. Here we briefly review a few of those most relevant to early English constructions.

Spatial relations

Constructions related to the domain of spatial relations are among the most commonly grammaticized and earliest acquired crosslinguistically, both in the single-word stage (*e.g.*, English *up*) and in the earliest word combinations. Semantic representations related to this domain have thus figured prominently in the literature: the case grammars of the kind proposed by Fillmore (1968) typically include *theme*, *goal* and *source* as roles in the thematic hierarchy; Jackendoff's (1990)'s lexical conceptual semantics includes *GO* as a conceptual primitive; and Talmy (2000) offers a comprehensive catalogue of spatial relations and their many dimensions of linguistically marked variation. As noted in Section 3.1.1, these notions have been represented using *image schemas* (Johnson 1987; Lakoff & Johnson 1980), which in turn lend themselves very naturally to ECG formalization. The SPG (Source-Path-Goal) schema of Figure 3.4 and the Trajector-Landmark, Container and Path schemas in Figure 3.8(d) illustrate how some of the most common image schemas can be defined.

Languages vary in the formal devices they exploit to express spatial relations. As we have seen already, argument structure constructions (such as the Directed-Motion construction from Section 3.1.1) may evoke a path schema. In addition, verbs like *enter* and *leave* specifically denote motion along a path and thus directly convey image-schematic information.[7] English has several classes of constructions especially relevant to the spatial domain. We have already introduced ex-

[7]Languages display typological preferences for whether paths are expressed mainly by verbs or satellites. The formalism allows flexible means of capturing either alternative, though as yet has no explicit encoding of this typological preference.

amples for the specific class of path-related expressions; more generally, these expressions include the following:

Prepositions link a word form with a spatial relation schema, such as the Trajector-Landmark schema or one of its subcases. The INTO-CXN from Figure 3.13 has a meaning pole that is typed as an instance of the Into schema, which in turn evokes and asserts bindings on instances of the Container and SPG schemas. It is defined as a subclass of PATH-PREP(osition) in view of its connection to the SPG schema.

Prepositional phrases are structured mappings with two constituents, a preposition and a referring expression, where the latter furnishes a landmark for the former, as exemplified for path expressions by the PATH-PHRASE construction in Figure 3.13.

Particles are represented as a kind of truncated prepositional phrase, of which only the relation is expressed; the landmark is assumed to be inferrable in context. Figure 3.20 defines an IN-PARTICLE as a specific path-related particle that evokes the Into schema of Figure 3.11 along with a Referent bound to its landmark. (Note that the HOME construction of the *Harry ran home* example similarly expresses a path, though with different lexically specific constraints.)

```
construction IN-PARTICLE
  subcase of PATH-PARTICLE
  form
    orth ⟵ "in"
  meaning : SPG
    evokes Referent as r
    evokes Into as i
    i.lm ⟵ r
```

Figure 3.20. The IN-PARTICLE construction

Variations on these constructions can be defined to accommodate crosslinguistic diversity in how spatial (and other) relations are marked. An ADPOSITIONAL phrase construction, for example, may allow either prepositional or postpositional word order. For languages that employ morphological case marking to indicate spatial relationships, morphological constructions similar to the prepositional phrase but with word-internal form constraints (as well as possible phonological constraints) could be defined. The meaning domain could also be broadened to encompass other kinds of relations (such as temporal ordering and possession).

Context-dependent constructions

Many of the earliest childhood constructions refer to the current discourse or situational context. We can capture the relevant meanings by explicitly evoking the DiscourseSpace schema defined in Figure 3.12 and constraining it appropriately. Two simple examples are shown in Figure 3.21:

Speech act constructions provide explicit cues to the speaker's intent. The IMPERATIVE construction is a simple mapping between a specific intonational contour and speech act type.

Deictic pronouns are defined as pronouns that refer directly to participants in the current discourse space. The YOU construction, for example, is linked to the current addressee.

```
construction IMPERATIVE
  subcase of INTONATIONAL-CXN
  form : Intonational-Contour
    type ⟵ falling
  meaning
    evokes DiscourseSpace as ds
    ds.speechAct ⟵ command
```

```
construction YOU
  subcase of PRONOUN
  form : Word
    orth ⟵ "you"
  meaning
    evokes DiscourseSpace as ds
    self_m ⟷ ds.addressee
```

Figure 3.21. Contextually grounded constructions: The IMPERATIVE and YOU constructions both evoke and constrain the DiscourseSpace schema.

Scaling up

The domains illustrated so far are not intended to capture the full range of formal devices and meaningful distinctions available in English, let alone crosslinguistically; they aim merely to provide reasonable coverage of the early utterances most relevant to the learning model. Nonetheless, our interest in the learning model to be proposed rests in part on the assumption that its target grammar formalism can scale gracefully beyond these to more complex linguistic phenomena. To date, ECG has been extended to handle the formal and semantic demands of many (adult) constructions in several typologically distinct languages, including:

- *Morphological constructions:* Morphologically complex words can be represented as structured mappings with constituents, with word-internal analogues of ordering relations used to capture concatenative morphological constraints. ECG analyses have been proposed for English verbal suffixes, Russian and Georgian case marking, and Hebrew verbal morphology.

- *Argument structure constructions:* A broader system of constructions capturing English argument structure and argument structure alternations has been developed (Dodge In Preparation; Dodge & Lakoff 2005).

- *Referential expressions:* Modifications of the Referent schema have been used to capture a variety of English referential phenomena, including modifiers, measure phrases like *a cup of tea*, and quantifiers.

- *Predicational expressions:* These include directional specifiers (deictic particles in Mandarin and German), English aspectual constructions, and preliminary studies of conditionals and hypotheticals in English and Mandarin.

3.4 Summary

This chapter has provided a brief overview of Embodied Construction Grammar, a representational formalism that supports a construction-based approach to grammar and a simulation-based model of language understanding. Two primary formal structures have been described:

- schemas, capturing generalizations over experience in the domains of form or meaning; and

- constructions, capturing generalizations over form-meaning pairs.

Both structures are represented using a common formalism that facilitate the parameterization of simulations of the kind described in Section 2.3.3. Together they provide a formally precise means of representing constructions, in the symbolic, unified and gestalt sense laid out in Section 2.2.1. The discussion here has centered on the needs of the learning model, drawing examples mainly from early English constructions. The schema and construction formalisms are, however, sufficiently expressive to describe a wide range of adult and crosslinguistic constructions. More importantly, they provide the crucial information needed for driving the dynamic language understanding processes to which we turn next.

Chapter 4

Using embodied constructions

Language is the source of misunderstandings.
— *Antoine de Saint-Exupéry*

We turn now from structure to process: the ECG formalism provides a means of expressing a variety of conceptual and linguistic relationships, but how do these static structures support dynamic processes of language use? This chapter takes a closer look at those processes, expanding on the informal account in Section 3.1.1 of the journey from sentence to simulation. Figure 4.1 updates the depiction in Figure 3.1 to distinguish three main processes involved in understanding an utterance in context:

- **Constructional analysis** identifies which constructions are instantiated by an utterance and how they are related.
- **Contextual resolution** links constructional meanings with the concrete entities and actions accessible in the situational context.
- **Embodied simulation** performs dynamic simulations that combine linguistically specified scenarios, world and embodied knowledge, and contextual constraints to yield new inferences about the situation.

A full-fledged model of language comprehension capturing the complexities of each of these processes, as well as the fine-grained interactions among them, is a long-term goal of the research

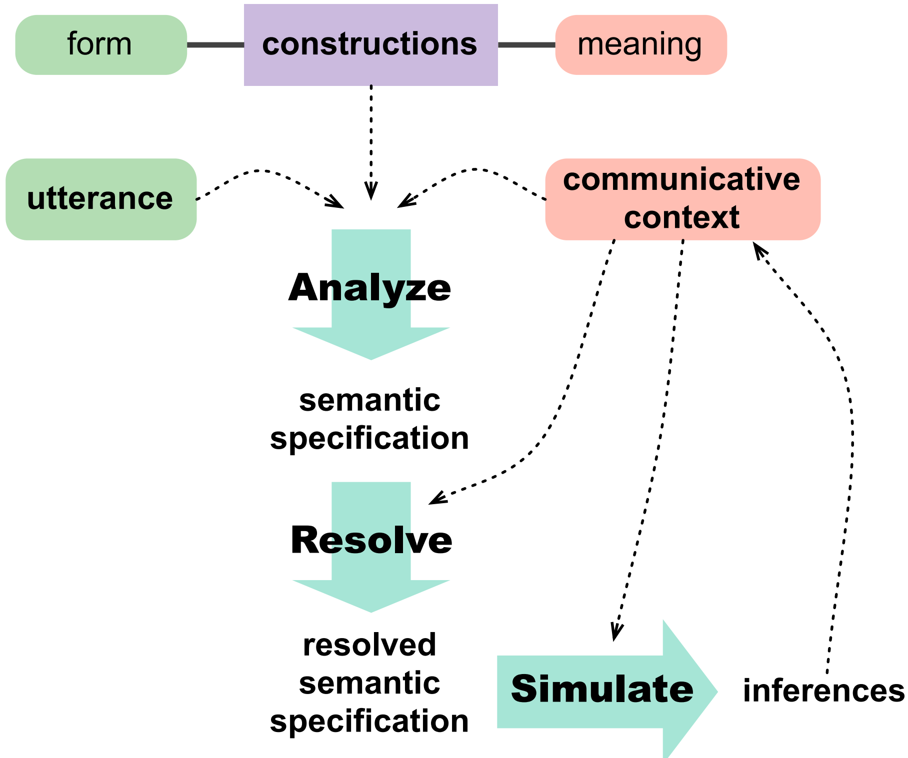

Figure 4.1. The main processes of the language understanding model: analyzing utterances in context, resolving context-dependent references and running embodied simulations.

project outlined in Section 1.2.2. In line with the present focus on early construction learning, this chapter takes on the more immediately tractable goal of providing simple models of the first two processes, analysis and resolution. Both of these processes may involve considerably more complications than in our introductory example. In general, an utterance may yield multiple candidate constructional analyses, and there may be multiple ways of resolving its referents in context. Each process thus requires both a means of identifying viable candidates and a means of evaluating those candidates to select the best one. Crucially, all of these mechanisms are directly tied to the learning model to be presented.

The chapter is organized as follows. Section 4.1 defines the structures and processes involved in analysis and resolution, focusing on the general problem constraints. Section 4.2 presents a simple integrated analyzer-resolver model suitable for use with the learning model under consideration, with evaluation strategies for both processes in Section 4.3.

4.1 Definitions

The basic formal structures underlying language understanding in the ECG framework — schemas and constructions — have already been introduced in Chapter 3. As depicted in Figure 4.1, the analysis process also requires input structures representing a specific utterance and its communicative context, and it produces additional structures, including the *semspec* that sets parameters for the simulation. This section defines all of these structures more precisely, in terms of their relation to schemas and constructions as well as their roles in the analysis and resolution processes.

4.1.1 Schemas and constructions

Grammar. We first revisit the ECG formalisms from Chapter 3 to highlight several aspects of both schemas and constructions that are especially relevant to the analysis and resolution processes. Both of these processes depend, first of all, on a well-defined means of comparing elements in both the schema and construction type hierarchies. Type relations are needed, for example, to find appropriate constituents for a construction or referents in context. The relevant relation is notated in the formalism by the use of the **subcase of** tag to specify immediate parents; these type relations are defined more precisely as follows:[1]

A **grammar** G is a pair (S, C), where S is a **schema set** and C is a **construction set** defined with respect to S.

- Each of S and C is a partially ordered set with a unique top element $\top$ and a well-defined $\mathrm{parent}(x, y)$ relation.
- For each element x the set of parents of x is denoted $\mathrm{parents}(x)$. That is, $\mathrm{parent}(x, y)$ if $x \in \mathrm{parents}(y)$.
- For each x, the transitive closure of the parent relation is denoted $\mathrm{ancestors}(x)$. In other words, x is an ancestor of y if y is a parent of a parent of $\ldots x$. We denote the relation "x is an ancestor of y" as $x \leq y$. This relation can also be stated as "y is a descendent of x"; that is, for each y, $\mathrm{descendents}(y)$ is the set of all x such that $x \leq y$.

The definitions below restate the key components of schemas and constructions and illustrate the parallels and connections between the two. In particular, while both (optionally) allow complex subordinate structure (through roles and constituents, respectively), constructions also have a more complex bipolar structure pairing schemas in their form and meaning poles.

[1] These definitions are extended in Section 6.3.1 to facilitate the cross-domain structural alignment of constructions.

Each **schema** $\mathbf{s}$ in schema set S is associated with:

- a name $\mathrm{name}(\mathbf{s})$;
- a domain $\mathrm{domain}(\mathbf{s}) \in \{\text{form}, \text{meaning}\}$;
- a set of parent schemas $\mathrm{parents}(\mathbf{s}) \subset S$;
- a set of local roles $\mathrm{roles}_{\mathrm{loc}}(\mathbf{s})$, where each role $r \in \mathrm{roles}_{\mathrm{loc}}(\mathbf{s})$ has a name $\mathrm{name}(r)$ and type $\mathrm{type}(r) \in S$ and may be marked as *evoked*; and
- a set of local constraints $\mathrm{constraints}_{\mathrm{loc}}(\mathbf{s})$, where each constraint $t \in \mathrm{constraints}_{\mathrm{loc}}(\mathbf{s})$ has a constraint type $\mathrm{type}(t)$.

Each schema $\mathbf{s}$ also inherits $\mathrm{roles}_{\mathrm{inh}}(\mathbf{s})$ and $\mathrm{constraints}_{\mathrm{inh}}(\mathbf{s})$ from its ancestors; $\mathrm{roles}(\mathbf{s})$ and $\mathrm{constraints}(\mathbf{s})$ are the union of the local and inherited roles and constraints, respectively.

Each **construction** $\mathbf{c}$ in construction set C (where C is defined with respect to schema set S) is associated with:

- a name $\mathrm{name}(\mathbf{c})$;
- a set of parent constructions $\mathrm{parents}(\mathbf{c}) \subset C$;
- a pair of **domain poles** ($\mathbf{c}_f$,$\mathbf{c}_m$), corresponding to the form and meaning poles, respectively, where each domain pole has a (form or meaning) type $t \in S$ and may add local roles and constraints; and
- a set of local constituent constructions $\mathrm{constituents}_{\mathrm{loc}}(\mathbf{c})$, where each $n \in \mathrm{constituents}_{\mathrm{loc}}(\mathbf{c})$ has a name $\mathrm{name}(n)$ and type $\mathrm{type}(n) \in C$, as well as constituent domain poles n_f and n_m.

Each construction $\mathbf{c}$ also inherits $\mathrm{constituents}_{\mathrm{inh}}(\mathbf{c})$, $\mathrm{roles}_{\mathrm{inh}}(\mathbf{c})$ and $\mathrm{constraints}_{\mathrm{inh}}(\mathbf{c})$ from its ancestors; $\mathrm{constituents}(\mathbf{c})$, $\mathrm{roles}(\mathbf{c})$ and $\mathrm{constraints}(\mathbf{c})$ are the union of the local and inherited constituents, roles and constraints, respectively.

That is, a construction $\mathbf{c}$ is a pairing of two graphs $\mathbf{c}_f$ and $\mathbf{c}_m$ (corresponding respectively to the form and meaning domains), each of which may have any number of elements and constraints. These domains may also be linked by the construction's constituents $\mathrm{constituents}(\mathbf{c})$; the informal distinction made in Section 3.3.1 between *simple* and *structured* constructional mappings can be restated in these terms as contingent on whether $\mathrm{constituents}(\mathbf{c})$ is empty (simple) or not (structured).

The potential for structural complexity in both domain graphs and constructional constituency gives rise to a range of possible mapping types, as illustrated in Figure 4.2. Figure 4.2(a-c) all depict simple mappings with no constructional constituents; they vary in how much structure is present in the form and domain domains, *i.e.*, whether $\mathbf{c}_f$ and $\mathbf{c}_m$ are represented as a single node. Canonical examples in English for these could include: (a) simple names for things, like *dog* or *Snuffleupagus*; (b) relational terms and verbs like *up*, *push*, *hi* (simple form, structured meaning) or fixed expressions for simple referents like *Mr. Rogers* (structured form, simple meaning); and (c) frozen expressions like *by and large*. The configuration in Figure 4.2(d), in contrast, has complex structure in both domains, corresponding to the structured, relational constructions that characterize most of the rest of language.

Crucially, these mapping categories are not fixed for a given expression, but rather depend on how a given grammar is defined. In fact, the same expression could be analyzed in terms of any

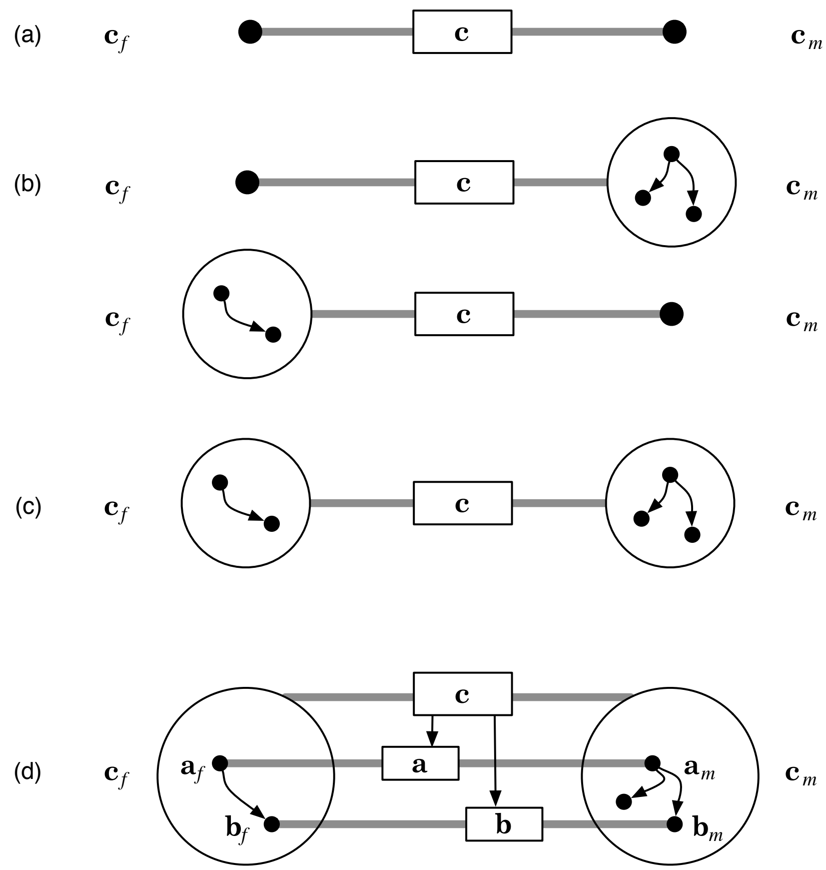

Figure 4.2. Cross-domain mapping types: (a) simple mapping between two entities; (b) simple mapping between an entity and a structured domain; (c) simple mapping between two structured domains; (d) structured relational mapping between two structured domains.

(or all) of the above types. A name like *Mr. Potato Head*, for example, could be represented as a fixed expression referring to a particular toy, where both domains are single-node gestalts; alternatively, the complex structure in each domain (words in sequence, or relations among a particular social status, species description and body part) could be represented, either with or without structured cross-domain mappings. The same variety of mappings is possible for apparently simple words, like *walking*, which could (in theory) be treated either as an unanalyzed morpheme or as a morphologically complex inflected verb. This representational flexibility is especially relevant to the learning context, in which the apprehension of increasingly complex structure in all domains allows the possible and preferred analysis of the same expressions to evolve over time.[2]

Types and tokens. It is important to distinguish between *types* and *tokens*. Both schemas and

[2]The fluidity with which meanings can be collapsed to a unitary node or expanded into an intricately structured web is also characteristic of many well-known semantic and pragmatic challenges. Within the broader simulation-based model of language understanding, different *construals* of a situation may correspond to different simulation conditions.

constructions are types, of which any particular instance encountered is a token. We refer to these tokens more specifically as **schema instances** and **constructs**, respectively:

A **schema instance** s instantiates a schema $\mathbf{s}$, where:

- $\mathrm{type}(s) = \mathrm{name}(\mathbf{s})$;
- s has an identifier $\mathrm{id}(s)$; and
- each role $r \in \mathrm{roles}(\mathbf{s})$ is a function on s to its filler schema instance $s.r$, where $\mathrm{type}(s.r) \leq \mathrm{type}(\mathbf{s}.r)$. The set $\mathrm{roles}(s)$ contains these schema instances $s.r$.

A **construct** c instantiates a construction $\mathbf{c}$, where:

- $\mathrm{type}(c) = \mathrm{name}(\mathbf{c})$;
- c has an identifier $\mathrm{id}(c)$; and
- c is associated with a pair of domain poles (c_f,c_m) isomorphic to ($\mathbf{c}_f$,$\mathbf{c}_m$), where $\mathrm{type}(c_f) \leq \mathrm{type}(\mathbf{c}_f)$ and $\mathrm{type}(c_m) \leq \mathrm{type}(\mathbf{c}_m)$; and
- each constituent $n \in \mathrm{constituents}(\mathbf{c})$ is a function on c to its filler construct $c.n$, where $\mathrm{type}(c.n) \leq \mathrm{type}(\mathbf{c}.n)$. The set $\mathrm{constituents}(c)$ contains these constituent constructs $c.n$.

A token x of a structure $\mathbf{x}$ is also notated $\mathrm{id}(x) : \mathrm{type}(x)$.

These notations specify that both schema instances and constructs derive their structure from the schemas and constructions they instantiate: their constituent schema instances and constructs stand in isomorphic relation to roles and constructional constituents, respectively; and constructs have form and meaning poles just as constructions do. The type constraints above require roles, constituent constructions and domain poles to be at least as specific as required by the defining schema or construction.

Usage statistics. The current framework assumes that both language understanding and language learning are sensitive to statistical regularities of various kinds. With respect to linguistic knowledge, such regularities include, minimally, the relative frequency with which constructions (or, more precisely, constructs) occur and co-occur. Below are definitions for several statistical measures, including the **count** of a construction $\mathbf{c}$ (its empirical frequency, *i.e.*, the sum over all processed tokens d of the number of times an instance of $\mathbf{c}$ appears in d's analysis); and its **weight** (its relative frequency, normalized over all the constructions in the grammar). These statistics are intended to capture constructional usage patterns defined history of data to which it has been exposed; analogous measures could be defined for a subcorpus of data D.

Constructions are associated with several measures of frequency:

- $\mathrm{count}(\mathbf{c})$, the raw frequency of $\mathbf{c}$, also notated $[\mathbf{c}]$
- $\mathrm{weight}(\mathbf{c})$, the relative frequency of $\mathbf{c}$, normalized over all the constructions in the grammar:

$$\mathrm{weight}(\mathbf{c}) = \frac{[\mathbf{c}]}{\sum_{\mathbf{x} \in G} [\mathbf{x}]}$$

- $\mathrm{ccount}(\mathbf{c}.n, \mathbf{x})$, the **constituency count**: the number of times construction $\mathbf{c}$ has its constituent n filled by construction $\mathbf{x}$, also notated $[\mathbf{c}.n \leftarrow \mathbf{x}]$
- $\mathrm{cweight}(\mathbf{c}.n, \mathbf{x})$, the **constituency weight**: the relative frequency of a constituent-filler pair, normalized over instances of $\mathbf{c}$:

$$\mathrm{cweight}(\mathbf{c}.n, \mathbf{x}) = \frac{[\mathbf{c}.n \leftarrow \mathbf{x}]}{[\mathbf{c}]}$$

These statistics can be incorporated into evaluation metrics for choosing among competing constructional analyses, as discussed in Section 4.3. The weight is also a reasonable estimate of a construction's prior probability; it thus also plays a role in the evaluation of competing grammars during learning, as discussed in Chapter 7.[3]

Examples. Figure 4.3 shows a mini-grammar G_1, containing several example schemas and constructions written using the ECG notation of Chapter 3. These include lexical constructions and associated meaning schemas for the words *ball* and *throw*, along with the lexically specific THROW-IMP construction that could license an imperative speech act used by a speaker to express a desire for the discourse addressee to undertake a particular throwing action. We can use the terms above to make some precise statements about these structures. For example, we can refer to the number of roles, constraints or constituents of a given structure: $|\,\mathrm{roles}(\mathsf{Ball})| = 2$ corresponds to the assertion that the Ball schema has two roles, and $|\,\mathrm{constituents}(\text{BALL-CN})| = 0$ asserts the lack of constituents for the lexical BALL-CN construction. Similarly, $\mathrm{type}(\text{THROW-IMP}.\mathsf{y})$ refers to the type of the THROW-IMP construction's y constituent — in this case, REF-EXPR.

4.1.2 Utterances and contexts

Both language understanding and language learning are situated within an ongoing stream of potentially relevant events, relations and discourse events. In general, these processes may exploit an extended situational and discourse history; for current purposes, I assume the stream of experience can be discretized into a **corpus** of **input token**s, each pairing an **utterance** with its communicative **context**, where the context serves as a static snapshot of the situation.

[3]Analogous measures could be introduced to capture usage statistics for the domains of form and meaning; see Bryant (2008) for a more complete probabilistic version of the formalism.

schema Ball
subcase of Toy
roles
color : Color
size : Size

construction BALL-CN
form
orth ⟵ "ball"
meaning : Ball

schema Throw
subcase of CausedMotion
roles
thrower : Animate
throwee : Object
constraints
action ⟵ THROW-XSCHEMA

construction THROW-CN
form
orth ⟵ "throw"
meaning : Throw

construction THROW-IMP
constituents
t : THROW-CN
y : REF-EXPR
form
t_f `before` y_f
meaning
evokes DiscourseSpace **as** ds
t_m.thrower ⟷ ds.addressee
t_m.throwee ⟷ y_m

Figure 4.3. ECG example schemas and constructions: schemas Ball and Throw, lexical constructions BALL-CN and THROW-CN, and imperative construction THROW-IMP.

A **corpus** D is a list of **input token**s d defined with respect to a schema set S. Each input token d is a pair (u, z), where u is an **utterance** and z is a communicative **context**.

Each **utterance** u is a feature structure associated with a set of form features, typically including:

- orthographic information (*e.g.*, a textual string);
- phonological information (*e.g.*, a phonological string); and
- intonational information (*e.g.*, a feature indicating the utterance's associated intonational contour).

The **context** is a feature structure associated with a set of context features, typically including:

- a set of participants;
- a set of schemas in the currently salient scene; and
- the current discourse space,

where all participants and schemas, along with the discourse space schema, are context items that instantiate schemas in S. Schema instances instantiated in context are also referred to as **context item**s.

Example. Figure 4.4 shows an example input token, represented as a pair of feature structures. The utterance is described with the text string "throw the ball" and a falling intonation; the accompanying context describes a scene of play between a mother and child (Naomi).[4] The object of

[4]For expositional clarity, context items corresponding to unique individuals (such as Mother and Naomi) are shown with only their associated schema names, while those that are instances of a larger category of items also include an instance identifier (such as bZ:Ball and dZ:Doll). By convention, context item identifiers end in Z, and construct identifiers end in C; schema instance identifiers sometimes (though not always) end in S. Atomic values are shown as lower-case text strings.

their current joint attention is a ball, which is also located proximally to the speaker. (The Trajector-Landmark and DiscourseSpace context items instantiate schemas defined in Chapter 3.)

$$
\left(
\begin{bmatrix}
\text{Utterance} & \\
\text{text}: & \text{"throw the ball"} \\
\text{intonation}: & \text{falling}
\end{bmatrix},
\begin{bmatrix}
\text{Context} & \\
\text{participants}: & \text{Mother, Naomi, bZ:Ball, dZ:Doll} \\
\text{scene}: & \begin{bmatrix} \text{tlZ:Trajector-Landmark} & \\ \text{trajector}: & \text{bZ:Ball} \\ \text{landmark}: & \text{Mother} \\ \text{relation}: & \text{proximal} \end{bmatrix} \\
\text{discourse}: & \begin{bmatrix} \text{dsZ:DiscourseSpace} & \\ \text{speaker}: & \text{Mother} \\ \text{addressee}: & \text{Naomi} \\ \text{speech act}: & \text{imperative} \\ \text{joint attention}: & \text{bZ:Ball} \\ \text{activity}: & \text{play} \\ \text{temporality}: & \text{ongoing} \end{bmatrix}
\end{bmatrix}
\right)
$$

Figure 4.4. A sample input token pairing the utterance "throw the ball" with a communicative context of mother-child play.

4.1.3 Analysis and resolution

We now turn to the analysis and resolution processes. It is useful to idealize these processes as running serially, as depicted in Figure 4.1: the output of constructional analysis is (part of) the input to contextual resolution. In practice, of course, these processes may run in parallel or interleaved fashion, and specific implementations may exploit their mutual constraints as heuristics for optimizing the search or evaluation procedures. Nonetheless, these processes remain functionally distinct, as defined by their associated input and output structures.

Analysis. The **constructional analysis** process is the analogue of syntactic parsing in standard grammatical formalisms: both aim to identify the grammatical structure of a sentence. With constructions, however, grammatical structure entails structure in both the form and meaning domains, as evident in the depiction of the analysis of *Harry ran home* shown earlier in Figure 3.3. That is, the **analysis graph** (or **construct graph**) consists of the constructs shown in the center of the diagram; this graph corresponds to a syntactic parse tree, but each construct of the graph has both form and meaning correlates. Taken together, the meanings associated with these constructs comprise the **semantic specification**. Constructional analysis thus performs some of the functions associated with semantic interpretation, generating something akin to a logical form or assignment of thematic roles. These two main results of the analysis process are defined below.

An **analysis** a_u of an utterance u, or a_d of an input token $d = (u, z)$, is a graph with nodes $\mathrm{constructs}(a)$ and edges $\mathrm{constituents}(a)$ corresponding to constituency relations on $\mathrm{constructs}(a)$.

- The set $\mathrm{roots}(a) \subseteq \mathrm{constructs}(a)$ contains all constructs of a that are not constituents of any other construct in a.
- a has associated domain poles (a_f,a_m), where a_f is the union of the form poles c_f and a_m is the union of the meaning poles c_m of all constructs $c \in \mathrm{constructs}(a)$.
- An analysis a is **partial** if it does not match all the forms in the utterance, that is, if $a_f \subset u$. a is **spanning** if a_f includes the initial and final units of u.

A **semantic specification** $\mathrm{semspec}(a_d)$ associated with an analysis a_d is the union of all meanings associated with a_d; that is, $\mathrm{semspec}(a_d) = a_m$. It is also denoted ss_a.

Resolution. The **contextual resolution** process grounds the interpretation produced by constructional analysis (*i.e.*, the semspec) with respect to the situational context. In contrast to the artificially context-free examples typical in syntactic and semantic analysis, all naturally occurring utterances are rooted within some ongoing discourse and situational context and must be interpreted accordingly. Even newspaper articles and other genres with relatively neutral contexts carry background assumptions about the (temporally and spatially displaced) interlocutors that affect the interpretation of, for example, indexicals (*yesterday, here*). The context-dependent nature of interpretation is especially pronounced in circumstances involving a high incidence of pronouns and situational referents, as in parent-child interactions.

Contextual resolution generalizes the standard notion of reference resolution beyond referring expressions to include all meanings evoked by an utterance. More precisely, the goal of the resolution process is to link the semspec produced by analysis with the current context. Typically, the semspec includes schema instances and role-fillers that are unspecific (*e.g.*, an instance of the category Human, or the rate parameters of a Walk schema) or make explicit reference to the current discourse space (*e.g.*, in the case of pronouns and other discourse-dependent constructions). Ideally, these schema instances should be linked, either directly or indirectly with contextually grounded schema instances (*i.e.*, context items), as expressed by a **resolution map**. Each resolution map represents a hypothesis about identity relations that hold between elements of the semspec and context.

A **resolution map** $r : ss_a \rightarrow z$ is a function from a semspec ss_a to a context z.

- Each relation $r(x) = y$ is a binding, also denoted $xis \Rightarrow y$, mapping a schema instance $x \in ss_a$ to a context item $y \in z$.
- The **resolved semspec** rss_r is the result of enforcing the bindings in r on $ss_a \cup z$.
- The set $\mathrm{mapped}_r(ss)$ is the subgraph of ss that is mapped by r onto z, *i.e.*, the subgraph whose root nodes are the domain of r.
- A resolution map r map is **total** if every element in ss_a is mapped to z, and **partial** otherwise.
- A resolution map is **valid** if each of its bindings successfully unifies.

Example. The results of analysis and resolution can be illustrated with an example from the child language learning domain, based on the mini-ECG grammar G_1 from Figure 4.3 and the input token based on *throw the ball* from Figure 4.4. That is, the (incomplete) input grammar contains the THROW, BALL and THROW-IMP constructions, but it lacks a construction for the determiner *the*. The resulting constructional analysis is depicted in Figure 4.5 (with some token details omitted for clarity). Note that the central analysis graph by itself resembles a traditional parse tree indicating hierarchical structure, *i.e.*, the constituency relations between the THROW-IMP construction and the two lexical constructions. It is also connected, however, with its matched forms (shown with heavy outline on the left) and associated semspec (on the right).

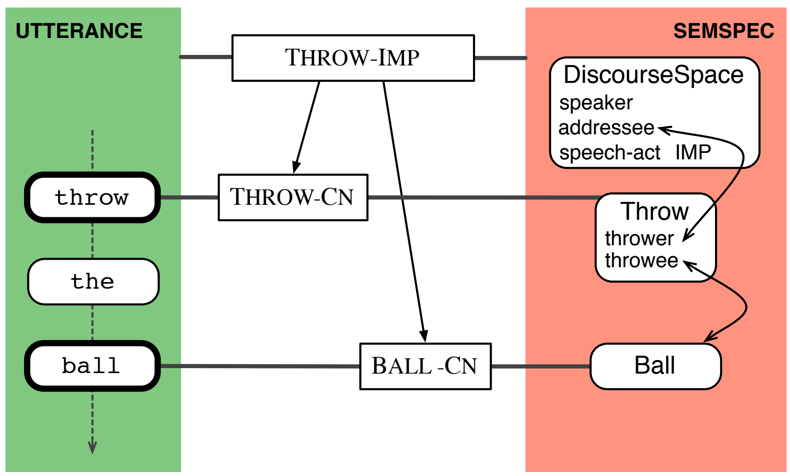

Figure 4.5. Analysis of an example input token, showing the analysis graph of constructs (center) linking utterances (left column) with a semspec (right column).

Continuing with this example, Figure 4.6 shows the result of contextual resolution: the semspec ball b:Ball, corresponding to the meaning pole bC_m of the bC:BALL-CN construct, is mapped to the specific ball bZ:Ball in context; and the discourse space ds:DiscourseSpace evoked by the THROW-IMP construction is mapped to the contextual discourse space dsZ:DiscourseSpace. These mappings, depicted as the two arrows linking the semspec and context representations (both simplified for clarity), constitute the resolution map.

The terms introduced above allow a more precise statement of all of these relations:

- constructs(a) = { tiC:THROW-IMP, tC:THROW-CN, bC:BALL-CN };
- constituents(a) = { tiC.t = tC, tiC.y = bC }, where $x.a = y$ denotes an edge from construct x based on constituent a to construct y;
- semspec(a_d) = {b:Ball, t:Throw, ds:DiscourseSpace }, where b = bC_m, t = tC_m and ds = tiC_m.ds; and
- r = {b $\Rightarrow$ bZ, ds $\Rightarrow$ dsZ}

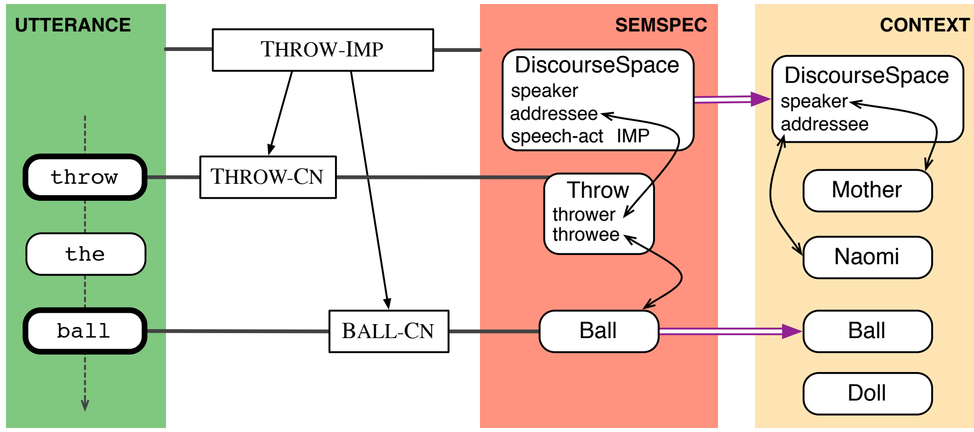

Figure 4.6. Analysis and resolution of example input token, showing the analysis graph of constructs linking utterances (left column) with a semspec (center column). The semspec is connected by resolution bindings (shown as double-lined arrows) to the current context (right column).

Note that the resolution map r includes no binding for the semspec item t:Throw, since the current context does not include any appropriate referents. This schema instance is nonetheless connected to contextual items indirectly, through its role bindings to other resolved semspec elements. (See Figure 4.7 and Figure 4.8 for the full feature structures corresponding to the semspec before and after resolution.)

* * *

I have now defined all the structures relevant to the language understanding processes at hand: linguistic knowledge in the form of an ECG grammar; a particular utterance to be interpreted and its immediate situational context, represented as an input token; the grammatical structure (an analysis graph) and semantic structure (the semspec) of an analyzed input token; and a mapping of semantic structure to situational context (a resolution map). Together these structures specify the inputs and outputs for the processes of constructional analysis and contextual resolution to be addressed next.

4.2 A simple analyzer-resolver

This section describes an integrated analyzer-resolver algorithm that satisfies both the input-output specifications laid out in the previous section and the broader needs of the language learning model to be described in Chapter 5. The examples in Section 3.1.1 and Section 4.1 involve tiny grammars that conveniently yield a single analysis, obviating the need to evaluate competing analyses; the relatively constrained set of contextually accessible referents makes resolution similarly straightforward. More realistic scenarios, however, afford a much larger space of possibilities for each of these processes.

The utterances in (4–1), taken from the Nomi corpus of the CHILDES database (Sachs 1983), exemplify the structural and referential ambiguity found in child-directed speech:

(4–1) a. Get the green one over there. *(referent = green puzzle piece)*
b. Should we put your pajamas on Nomi?

Each of these examples yields multiple candidate constructional analyses, and (depending on its associated situation) may have multiple ways of resolving its contextual referents. In (4–1a), the potential for structural and referential ambiguity hinges on whether the phrase *over there* serves as a post-nominal modifier within a larger referring expressing indicating which green item to retrieve (*the green one over there*), or specifies the goal of a requested caused motion. These alternative bracketings, shown in (4–2), entail different constraints on the search for the particular item to be retrieved or moved.

(4–2) Get the green one over there.
a. [Get [the [green one] [over there]]] . *(retrieval: distal referent, proximal goal)*
(≈ "(You) retrieve the green one that is over there.")
b. [Get [the [green one]] [over there]] . *(caused motion: distal goal)*
(≈ "(You) cause the green one to move over there.")

Likewise, in (4–2b), *Nomi* may be interpreted as an utterance-final tag (naming the addressee), or as the landmark of the spatial phrase *on Nomi* (specifying the intended pajama-wearer).

As these examples demonstrate, the two processes, while conceptually distinct, are interdependent; their joint goal is to find the best mutually consistent solution that satisfies their collective constraints. The domain of child-directed speech imposes some further requirements: incomplete grammars are the norm rather than the exception, so analyses are typically partial; and the primacy of the immediate situational context permits some relaxation of the referential conventions of adult-directed speech. All of these considerations suggest that both processes should provide a means of determining not just *whether* a given candidate solution (*i.e.*, an analysis or resolution map) is viable but also *how well* it accounts for the input.

A high-level integrated analysis-resolution algorithm meeting these requirements is summarized below:

Analyze and resolve utterance in context: Given an input token $d = (u, z)$ and a grammar G, return the best analysis a_d and its associated resolution map r_a.

1. Analyze utterance u using grammar G, producing a set of candidate analyses A.
2. For each candidate analysis $a \in A$:
   (a) Resolve $ss_a = \mathrm{semspec}(a)$ against context z, producing a set of candidate resolution maps R_a.
   (b) Set r_a = the resolution map $r \in R_a$ that maximizes the resolution score $\mathrm{score}(ss_a, r, z)$.
3. Return a_d = the analysis $a \in A$ and associated resolution map r_a that maximize $\mathrm{score}(a, r_a, d)$.
4. Reinforce each construction $\mathbf{c}$ instantiated in a_d by incrementing its count $\mathrm{count}(\mathbf{c})$; adjust grammar weights.

This algorithm casts the combined processes of analysis and resolution as a nested search over their respective domains. The sections to follow describe implementations of each of the search strategies, as well as criteria for evaluating their resulting candidates. All aspects of these processes admit a variety of implementations; the guiding motivation here is to employ the simplest strategies that fulfill the functional criteria of the language learning model — *i.e.*, that provide robust semantic interpretations of utterances in context, based on the input and output structures specified in Section 4.1. As noted earlier, these structures are, by and large, cognitively motivated approximations of the abilities (and limitations) of a child learner. The specific implementational details of the processes below are not, however, intended as a realistic psycholinguistic model of language understanding. (See Bryant (2008) for recent work on a cognitively more plausible construction analyzer that incrementally combines constructional, syntactic and semantic cues to find the best-fitting analysis of an utterance.)

4.2.1 Constructional analysis

The goal of the constructional analysis process is to piece together one or more sets of constructions to account for the input. This task is subject to many of the same challenges faced by syntactic parsers; these include, most prominently, the aforementioned issues of structural ambiguity and incomplete grammars, along with the potentially large search space of analyses that may render an exhaustive search procedure intractable. These challenges are further compounded by the inclusion of meaning, since the constructional meanings associated with an analysis must be unified in a coherent semspec.

The construction analyzer used here adapts many techniques from syntactic parsing to exploit semantic information, as well as general search techniques for prioritizing promising analyses, pruning unpromising analyses and preventing redundant processing. As described in more detail in Bryant (2003), the analyzer combines the partial parsing (or *chunk* parsing) approach taken by Abney (1996) with a unification-based parsing strategy for satisfying constraints from multiple domains. It also employs a *chart* for recording intermediate results and eliminating redundant processing. The resulting analyzer is especially suited to processing data using incomplete grammars: even a small set of simple constructions can support skeletal interpretations of complex utterances, a crucial requirement of the child language-learning scenario.

Like other partial parsers, the analyzer proceeds by identifying progressively larger chunks of licensed material; these islands of certainty (potentially amid a sea of unanalyzable material) serve as the basis of a partial analysis even when the grammar is incomplete. The analyzer consists of an array of **construction recognizers**, one for each construction in the input grammar; constructions are organized into successive levels, with lexical constructions at the lowest level (0) and more complex constructions at higher levels.[5]

Each recognizer attempts to match the input forms of its associated construction, in a manner analogous to Abney's (1996) finite-state recognizers. For simple (lexical) constructions, this process amounts to checking whether any of the input words matches the orthographic feature specified in its form pole. For structured constructions with constituents, the recognizer seeks the chart for candidate fillers for each of its constituents and checks whether its semantic constraints are satisfied by any suitable (sets of) candidates. If so, this successful construct is added to the chart. After all matching constructions have been found, the analyzer uses the chart to identify candidate analyses, consisting of non-overlapping subsets of the constructs in the chart. As in Bryant (2003), the analyzer restricts candidate analyses to spanning analyses with a minimum number of root constructs; this preference favors analyses that account for as much of the input utterance as possible.

The analysis algorithm is summarized below:

Analyze utterance: Given an utterance u and a grammar G, return a set of analyses A_u.

1. For each word w in position p of u, add a construct instantiating each matching lexical construction to chart at level 0 and position p.
2. For each construction level i from 1 to the highest level:
   (a) For each construction $\mathbf{c}$ and position p, run recognizer for $\mathbf{c}$ starting from p. Add construct instantiating $\mathbf{c}$ to chart at level i and position p if all constructional, form and meaning constraints are satisfied.
   (b) Repeat until no new constructions are recognized at level i.
3. Return the set A_u of spanning analyses a_u in the chart with a minimum number of roots.

[5]The assignment of constructions to levels is a convenient means of restricting recursion and avoiding the need to backtrack over alternate analyses, but it has no bearing of consequence on the score calculation or the learning model.

Evaluation criteria for analyses are described in Section 4.3; these include criteria for both unresolved analyses (as produced by the algorithm above) and resolved analyses (if the context is also available for contextual resolution).

Example. We return to the simple example from Section 4.1.3 to show the analyzer algorithm in action. (More complex cases involving multiple candidate analyses is described in Section 4.3.3.) The analyzer first seeks lexical matches to the input sentence and adds the corresponding constructs (tC:THROW-CN and bC:BALL-CN) to the chart. (No such construction is found for the word form "the".) Next, the analyzer attempts to match its non-lexical constructions; while running the recognizer for the THROW-IMP construction, it finds candidate fillers in the chart for the two constituents (while skipping the word form *the*); after succeeding in checking the specified form constraint (on their word order) and the various role-filler bindings asserted on their meanings, it adds the recognized construct tiC:THROW-IMP to the chart.

Figure 4.7 shows the resulting completed chart, along with the semspec from the best analysis; these correspond to the graphic depiction of Figure 4.5. In fact, this simplified example yields only one spanning analysis, with the single root tiC and constituents filled by lexical constructs tC and bC. (Another spanning analysis, consisting of only these lexical constructs is filtered out, since it has two roots.) The chart notation tiC:Throw-Imp [tC, bC; sk=1] serves as a shorthand for identifying a construct's constructional type and constituent fillers, as well as indicating how many forms, if any, are skipped. Root constructs of (the best) candidate analyses are shown in bold in the chart. The semspec shown here corresponds to that of (the sole) best analysis: like the context representation defined in Section 4.1, is is organized into its participants (a ball); its scene (an instance of the Throw schema; and its discourse information (an instance of DiscourseSpace). (The speaker and addressee are understood to be special participant roles in the discourse space.) The shared variables in various role-fillers indicate that the addressee is identified with the intended thrower, and the ball is identified with the intended throwee.

4.2.2 Contextual resolution

The goal of the contextual resolution process is to find one or more resolution maps linking a semspec to an input context. That is, the meanings evoked during constructional analysis must be grounded with respect to the current situational and discourse context.

This process involves many potential complications: there may be multiple candidate referents for a given expression (potentially ranked by salience, recency or other criteria); a referent may

tiC: Throw-Imp [tC, bC; sk=1]		
tC: Throw-Cn		bC: Ball-Cn
`throw`	`the`	`ball`

$$\begin{bmatrix} \text{Semspec} & \\ \text{participants}: & \text{b:Ball} \\ \text{scene}: & \begin{bmatrix} \text{t:Throw} & \\ \text{thrower}: & \text{ad} \\ \text{throwee}: & \text{b} \end{bmatrix} \\ \text{discourse}: & \begin{bmatrix} \text{ds:DiscourseSpace} & \\ \text{speaker}: & \text{sp:Human} \\ \text{addressee}: & \text{ad:Human} \\ \text{speech act}: & \text{imperative} \end{bmatrix} \end{bmatrix}$$

Figure 4.7. Constructional analysis: chart and semspec for the (partial) analysis of *throw the ball*, based on an incomplete input grammar. The best analysis, with root construct shown in bold, skips one form, the word *the*.

be available only indirectly, *i.e.*, conceptually related to an explicit referent but not present in the situation or mentioned in discourse; and subtle aspects of information structure may govern the particular ways in which different referents are likely to appear. Moreover, contextual resolution is intended to cover broader territory than the resolution of discourse anaphora or the determination of coindexation relationships among discourse referents. Particular constructions might, for example, effect the addition of new contextual items (*Harry baked a cake*) or interact with processes of metaphorical or aspectual construal (*Harry kicked the bucket*).

The parameters of the current investigation are, however, much more restricted. Child-directed speech differs from many of the genres traditionally addressed by previous work in either computational approaches to discourse modeling or linguistic studies of reference and discourse anaphora. In general, discourse history is subordinated to situational context: the vast majority of expressions make direct reference to the immediate environment, often overriding referential conventions of less context-bound genres. Anaphora frequently appear with no explicitly mentioned prior antecedent, referring instead to a situationally available entity. The discourse functions associated with definite and indefinite reference are tilted toward referring to entities already present in context; indefinite reference is typically used in a labeling context or to make non-specific reference to one of several items in a set. Deictic expressions are common: many utterances have explicitly deictic elements (*here, over there, this, that one*), often accompanied by non-linguistic cues.

The current work thus adopts a simple resolution strategy suitable for the needs of the learning model, essentially performing an exhaustive search over the space of resolution maps between a semspec and the context. No context history across utterances is assumed, and no explicit creation of referents (*e.g.*, based on indefinite referring expressions) is currently allowed. The algorithm below relies mainly on type-based and unification constraints: for each schema instance in the semspec, it first finds all potential referents in context, where the context item should be of a more specific type than the semspec item. It then generates all possible resolution maps, where each semspec item is bound either to one of its candidate referents or to no referent. Some of these resolution maps may include bindings that are either independently or collectively inconsistent, causing unification to fail; these are eliminated. The set of remaining valid maps with a maximum number of bindings (*i.e.*, ones that bind the most semspec items) is returned.

Resolve semspec: Given semspec ss and context z, return a set of resolution maps R.

1. Create a mapping of candidate bindings $cb : ss \rightarrow \mathcal{P}(z)$, where $cb(x) = \{y \in z : \text{type}(y) \leq \text{type}(x)\}$. That is, for each schema instance $x \in ss$, $cb(x)$ is the set of candidate context items that are type-descendants of x.
2. A candidate resolution map $r : ss \rightarrow z$ is a function from a subset of ss to z, such that for each $x \in ss$, $r(x) \in cb(x)$. Let R be the set of all candidate resolution maps r.
3. Remove invalid resolution maps from R. For each $r \in R$, create a resolved semspec rss_r:
   (a) Initialize rss_r as $ss \cup z$.
   (b) For each binding $x \Rightarrow y \in r$, unify x and y in rss_r. If unification fails for any binding, remove r from R and continue to next resolution map.
4. Return the set R of valid resolution maps with a maximum number of resolution bindings.

Example. We continue with the example from Section 4.1.3. Assuming the analysis and semspec in Figure 4.7, the resolver algorithm takes as input:

- ss = {b:Ball, t:Throw, ds:DiscourseSpace}
- z = {Mother, Naomi, bZ:Ball, dsZ:DiscourseSpace, tlZ:TrajectorLandmark }

and proceeds as follows:

1. cb(b) = {bZ}, cb(ds) = {dsZ}, cb(t) = {}
2. R = { $r1$ = { b ⇒ bZ, ds ⇒ dsZ}, $r2$ = {b ⇒ bZ}, $r3$ = {ds ⇒ dsZ}, $r4$ = { } }
3. No invalid maps found in R.
4. Return R = { $r1$ = { b ⇒ bZ, ds ⇒ dsZ} }

That is, the resolution process finds a single candidate binding for each of b and ds and no candidate bindings for t. Candidate resolution maps are created by choosing at most one binding from each $cb(x)$; this yields four resolution maps. Each of these maps is valid, in that the unifications they assert succeed; but only the largest resolution map ($r1$) is returned. Figure 4.8 shows the input semspec and context linked by this map, with resolution bindings indicated as the boxed coindexation labels $\boxed{1}$ and $\boxed{2}$. In standard unification fashion, these bindings recursively entail bindings on their respective subsidiary roles (as shown by the $\boxed{2\text{-}1}$, $\boxed{2\text{-}2}$ and $\boxed{2\text{-}3}$ indices on the respective DiscourseSpace roles).

$$
\left[\begin{array}{ll}
\text{Semspec} & \\
\text{participants}: & \text{b:Ball}\ \boxed{1} \\
\text{scene}: & \left[\begin{array}{ll} \text{t:Throw} & \\ \text{thrower}: & \text{ad} \\ \text{throwee}: & \text{b} \end{array}\right] \\
\text{discourse}: & \left[\begin{array}{ll} \text{ds:DiscourseSpace}\ \boxed{2} & \\ \text{speaker}: & \text{sp:Human}\ \boxed{2\text{-}1} \\ \text{addressee}: & \text{ad:Human}\ \boxed{2\text{-}2} \\ \text{speech act}: & \text{imperative}\ \boxed{2\text{-}3} \end{array}\right]
\end{array}\right]
$$

$$
\left[\begin{array}{ll}
\text{Context} & \\
\text{participants}: & \text{Mother, Naomi, bZ:Ball}\ \boxed{1}\text{, dZ:Doll} \\
\text{scene}: & \left[\begin{array}{ll} \text{tlZ:Trajector-Landmark} & \\ \text{trajector}: & \text{bZ:Ball} \\ \text{landmark}: & \text{Mother} \\ \text{relation}: & \text{proximal} \end{array}\right] \\
\text{discourse}: & \left[\begin{array}{ll} \text{dsZ:Discourse}\ \boxed{2} & \\ \text{speaker}: & \text{Mother}\ \boxed{2\text{-}1} \\ \text{addressee}: & \text{Naomi}\ \boxed{2\text{-}2} \\ \text{speech act}: & \text{imperative}\ \boxed{2\text{-}3} \\ \text{joint attention}: & \text{bZ:Ball} \\ \text{activity}: & \text{play} \\ \text{temporality}: & \text{ongoing} \end{array}\right]
\end{array}\right]
$$

Figure 4.8. Result of analyzing and resolving example input token pairing the utterance "throw the ball" with the specified context, using the resolution map $\{\text{b} \Rightarrow \text{bZ}, \text{ds} \Rightarrow \text{dsZ}\}$.

The resolved semspec resulting from unifying the structures bound by the resolution map is shown in Figure 4.9: the role-filler information available from the input context is now explicitly supplemented with connections to the semspec. Note that the explicit bindings asserted by the resolution map $\{\text{b} \Rightarrow \text{bZ}, \text{ds} \Rightarrow \text{dsZ}\}$ are not the only identifications present in the resolved semspec: the thrower and the discourse addressee roles are filled by the same participant (Naomi), as are the throwee and the discourse joint attention roles; these **inferred bindings** are entailed by the particular resolution map hypothesized here. Each of these context-semspec bindings, direct or indirect, serves to link a grounded situational participant, relation or event to the linguistically interpreted semspec, and thus to the meaning poles of constructs in the analysis graph. As we will see in Chapter 6, these potential links between language and the world directly guide the search for new constructional mappings.

$$
\left[\begin{array}{ll}
\text{Resolved semspec} & \\
\text{participants}: & \text{Mother, Naomi, b:Ball=bZ:Ball, dZ:Doll} \\
\text{scene}: & \left[\begin{array}{ll}\text{t:Throw} & \\ \text{thrower}: & \text{Naomi} \\ \text{throwee}: & \text{bZ:Ball}\end{array}\right], \left[\begin{array}{ll}\text{tlZ:Trajector-Landmark} & \\ \text{trajector}: & \text{bZ:Ball} \\ \text{landmark}: & \text{Mother} \\ \text{relation}: & \text{proximal}\end{array}\right] \\
\text{discourse}: & \left[\begin{array}{ll}\text{ds:DiscourseSpace=dsZ:DiscourseSpace} & \\ \text{speaker}: & \text{Mother} \\ \text{addressee}: & \text{Naomi} \\ \text{speech act}: & \text{imperative} \\ \text{joint attention}: & \text{bZ:Ball} \\ \text{activity}: & \text{play} \\ \text{temporality}: & \text{ongoing}\end{array}\right]
\end{array}\right]
$$

Figure 4.9. Resolved semspec resulting from analyzing example input token

4.3 Evaluation criteria

The integrated analyzer-resolver just described casts the language comprehension problem as two related search problems, one over the space of constructional analyses and the other over the space of resolution maps. These searches should yield the (resolved) analysis and semspec that together best account for the utterance in the current context — in other words, the best explanation of why a particular configuration of forms occurs with a particular configuration of meanings.

What makes one analysis or resolution map better than another? The stated goal of language comprehension within the current framework is to *make sense* of the input situation. To formalize this notion, it is useful to observe that these processes involve structures across three domains, mirroring the structure of a complex construction. Each of these domains may exhibit domain-specific regularities that make some configurations preferable to or more likely than others. The particular evaluation criteria most appropriate for these domains, both separately and in combination, depend on the goals and constraints of the modeling task.

The sections below define the various evaluation metrics used in the current model. These include scoring criteria for the constructional domain (Section 4.3.1) that favor more likely and more specific construct graphs, inspired largely by its syntactic parsing analogue; and for the form and meaning domains (Section 4.3.2) that favor more complete coverage of the respective input structures. These domain-specific criteria serve as the basis for the analysis score (Section 4.3.3) used to select among competing analyses (resolved or unresolved).

4.3.1 Scoring construct graphs

A construct graph is a grammar-dependent hypothesis about the network of constructions linking a particular input utterance and context. Considered independently of their domain poles, construct graphs can be evaluated only with respect to structural or statistical properties of their associated input grammar. Intuitively, graphs involving fewer, more frequent and more specific constructions that co-occur with high frequency, or that have shorter derivations, might be seen as more succinct and precise predictors of the input data.

The current model defines a **constructional score** score_c intended to encode this preference for predictive value, motivated by probabilistic approaches to the evaluation of syntactic parse trees. Recall that an analysis a is a set of constructs with constituency relations; each construct c (instantiating some construction **c**) is either a root construct or a constituent of another construct in the analysis.

We can approximate $P(a)$ the joint probability of the constructs in a as the product of the probabilities of its root constructs r and the conditional probabilities $P(c.n|c)$ for each construct c of its constituents $c.n$ given c:

$$\begin{aligned} P(a) &= P(c_1, c_2, \ldots c_m) \\ &= P(c_1) \times P(c_2|c_1) \times \ldots \times P(c_m|c_1, \ldots, c_{m-1}) \\ &\approx \prod_{r \in \mathrm{roots}(a)} P(r) \prod_{\substack{c \in a, \\ c.n \in \mathrm{constituents}(c)}} P(c.n|c) \end{aligned} \tag{4.3.1}$$

This term $P(c|c.n)$ accounts for the constituent-filler relationships in the analysis, that is, the probability that a construct c has its n constituent filled by its filler in this analysis $c.n$.[6] Equation (4.3.1) relies on some simplifying conditional independence assumptions: (1) the root constructs are conditionally independent of all other constructs; (2) non-root constructs are conditionally independent of all other constructs except those of which they are constituents; and (3) the constituent probabilities of a construct are conditionally independent.

The token-specific probabilities in (4.3.1) (*i.e.*, $P(r)$ and $P(c.n|c)$) can be further approximated using type-based probabilities; the usage statistics defined in Section 4.1.1 for ECG grammars provide convenient estimates for these terms:

$$P(r) \approx P(\mathbf{r}) \approx \mathrm{weight}(\mathbf{r}) \tag{4.3.2}$$

$$P(c.n|c) \approx P(\mathrm{type}(c.n)|\mathbf{c}, \mathrm{type}(\mathbf{c}.n)) \approx \mathrm{cweight}(\mathbf{c}.n, \mathrm{type}(c.n)) \tag{4.3.3}$$

[6] These conditional constituent probabilities are analogous to rule probabilities in stochastic context-free grammars.

where $\mathrm{weight}(\mathbf{r})$ is the relative frequency of the root constructions, $\mathrm{type}(c.n)$ is the observed type of the relevant constituent construct $c.n$, and $\mathrm{cweight}(\mathbf{c}.n, \mathrm{type}(c.n))$ is the constituency weight, defined in Section 4.1.1 as the relative frequency of a constituent $\mathbf{c}.n$ being filled by a construct whose type is $\mathrm{type}(c.n)$.

Under some circumstances, the requisite statistical information may be unavailable or unreliable. During early stages of child language learning, for example, most constructions are lexical and therefore have no constituents; statistics for many constituent-filler pairs may be quite sparse. The several options for the score_c metric below represent increasingly coarse approximations of $P(a)$.

Constituent probability. The approximations in (4.3.2) and (4.3.3) are used to define a *constructional score* of an analysis a, $\mathrm{score}_c(a)$:

$$\mathrm{score}_c(a) = \prod_{r \in \mathrm{roots}(a)} \mathrm{weight}(\mathbf{r}) \times \prod_{\substack{c \in a, \\ n \in \mathrm{constituents}(\mathbf{c})}} \mathrm{cweight}(\mathbf{c}.n, \mathrm{type}(c.n)) \tag{4.3.4}$$

$$= \prod_{r \in \mathrm{roots}(a)} \mathrm{weight}(\mathbf{r}) \times \prod_{\substack{c \in a, \\ n \in \mathrm{constituents}(\mathbf{c})}} \frac{[\mathbf{c}.n \leftarrow \mathrm{type}(c.n)]}{[\mathbf{c}]} \tag{4.3.5}$$

Uniform constituent probability. An alternative to using constituency weights is to assume a uniform probability over all possible constituent filler types. That is, each $\mathrm{cweight}(\mathbf{c}.n, \mathrm{type}(c.n))$ in (4.3.4) is set to the reciprocal of the number of construction types subsumed by $\mathrm{type}(\mathbf{c}.n)$ (*i.e.*, the type associated with role $\mathbf{c}.n$) in the grammar:

$$\mathrm{score}_{c_u}(a) = \prod_{r \in \mathrm{roots}(a)} \mathrm{weight}(\mathbf{r}) \times \prod_{\substack{c \in a, \\ n \in \mathrm{constituents}(\mathbf{c})}} \mathrm{cweight}_u(\mathbf{c}.n, \mathrm{type}(c.n)) \tag{4.3.6}$$

$$= \prod_{r \in \mathrm{roots}(a)} \mathrm{weight}(\mathbf{r}) \times \prod_{\substack{c \in a, \\ n \in \mathrm{constituents}(\mathbf{c})}} \frac{1}{|\,\mathrm{descendents}(\mathrm{type}(\mathbf{c}.n))|} \tag{4.3.7}$$

The constructional root prior probabilities expressed by $\mathrm{weight}(\mathbf{r}_a)$ may likewise be replaced by a uniform prior over all constructions in the grammar. That is, for all $\mathbf{c}$, $\mathrm{weight}(\mathbf{c}) = \frac{1}{|G|}$, where $|G|$ is the number of constructions in G.

Constructional compactness. Another alternative appropriate for settings in which statistical information is unavailable is to directly measure structural properties of the construction graph. The two most relevant measures are (1) the number of constructs in an analysis a, or $|\,\mathrm{constructs}(a)|$; and (2) the derivation length of a, or $\mathrm{depth}(a)$. Combined in (4.3.8) below, these provide an approximate measure of *compactness* that favors shorter, more specific analyses:[7]

$$\text{score}_{c_c}(a) = \frac{1}{|\,\text{constructs}(a)| + \text{depth}(a)} \tag{4.3.8}$$

The simple criteria given above for evaluating constructional graphs could be enriched to accommodate additional cognitive or processing considerations. The independence assumptions above could be relaxed to accommodate more complex interactions among constituents and across larger-scale discourse and situational contexts, as when constructions are primed by recent use. For the early language learning domain, however, the several $\text{score}_c(a)$ functions above should suffice.

4.3.2 Scoring form and meaning coverage

The form and meaning domains lend themselves to more direct evaluation: the results of the analysis and resolution processes should be maximally consistent with the input token. That is, the input utterance and context provide some basis against which the analyzed forms and resolved meanings may be compared. In general, these might not be expected to match exactly. The schematic forms and meanings associated with an analysis are, in fact, typically less well-specified than the concrete, grounded values in an input token: schemas by their nature abstract away from the particular phonological realization or stress pattern associated with a word form, or the actual color, size or material associated with some physical object.[8] Moreover, the meaning domain may exhibit an even greater discrepancy between the semspec and context, depending on the speech act involved. Imperatives (such as the *throw the ball* example) may specify a meaning that is not available in the context (such as a desired throwing event), even if it is partially grounded in context (in this case, since its roles are resolved to contextually accessible entities).[9]

Despite this potential for mismatch, for evaluation purposes it is still useful to idealize each input token d as providing a target for its (resolved) analysis a to match, with the goal of maximizing overlap between analyzed forms and meanings and their counterparts in the observed input. To measure this overlap, we draw on the *precision* and *recall* measures widely used in information retrieval and computational linguistics tasks:

$$\text{precision} = \frac{|a \cap d|}{|a|} \qquad\qquad \text{recall} = \frac{|a \cap d|}{|d|} \tag{4.3.9}$$

[7] The probabilistically inspired measures above also implicitly cover a notion of compactness. The preference for compactness anticipates the use of length-based evaluation criteria for the learning model, as motivated and defined in Chapter 7.

[8] In practice, input tokens available for a particular modeling task may also elide many such subtleties, allowing a closer match in each domain between the content predicted by an analysis and that observed in the input.

[9] Again, these discrepancies depend on input assumptions. As described in Chapter 5, the input tokens used for the learning task typically include schemas the child is likely to infer from context, such as the action specified by an imperative.

where precision measures how much of the (resolved) analysis is present in the input token, and recall measures how much the input token is present in the (resolved) analysis. Below I adapt and motivate these metrics for each of the two domains.

Form

The analysis algorithm given in Section 4.2.1 chooses only analyses consistent with the input utterance. Thus, for an analysis a of an input token $d = (u, z)$, $a_f \subseteq u$, which guarantees that precision in the form domain will be 100%. The form score used here is thus defined only in terms of recall:

$$\mathrm{score}_f(a, u) = \mathrm{recall}_f(a, u) = \frac{|a_f \cap u|}{|u|} \tag{4.3.10}$$

where each of the sets includes both form units and form relations, which may be weighted. The $\mathrm{score}_f(a, u)$ term measures how much of the utterance u an analysis accounts for.

Meaning

The meaning domain poses more challenges to evaluation. Informally, language comprehension should yield meanings that are easily interpretable in context. Within the broader language understanding model, analyses and their associated semspecs may be evaluated in terms of the simulations they parameterize, preferring those that yield the most relevant, informative inferences. For current purposes, we assess interpretability based on the degree of overlap between the (resolved) semspec and the context, or (for an unresolved semspec) structural properties of the semspec itself.

Resolved meaning. The overlap between semspec and context is directly captured by the resolution map resulting from the resolution process. For scoring purposes, however, we need to include not just the top-level schemas identified in the map, but also the substructures of these schemas that are recursively bound to contextually grounded items and the identification bindings among them — that is, the subgraph $\mathrm{mapped}_r(ss)$ of a semspec ss mapped by resolution map r. Semantic recall and precision are defined in terms of this mapped subset; their harmonic mean, or fscore, is used as the resolved meaning score score_{m_r}:

$$\mathrm{precision}_m(ss, r, z) = \frac{|\,\mathrm{mapped}_r(ss)|}{|ss|} \qquad \mathrm{recall}_m(ss, r, z) = \frac{|\,\mathrm{mapped}_r(ss)|}{|z|} \tag{4.3.11}$$

$$\mathrm{score}_{m_r}(ss, r, z) = \mathrm{fscore}(ss, r, z) = \frac{2 \cdot \mathrm{precision}_m(ss, r, z) \cdot \mathrm{recall}_m(ss, r, z)}{\mathrm{precision}_m(ss, r, z) + \mathrm{recall}_m(ss, r, z)} \tag{4.3.12}$$

That is, the precision measures how much of the linguistically analyzed meaning is resolvable in context, while the recall measures how much of the context is predicted by the linguistic analysis.

All of these values include both schemas and bindings: the size $|rss - z|$ measures not only the schemas in ss that are mapped in rss, but also the identification bindings among the roles of ss that hold of their mapped counterparts in rss.

Unresolved meaning. When contextual information is unavailable, an alternate meaning score based on properties of the unresolved semspec is more suitable; two candidate measures are (1) *value density* ($\mathrm{density}_v$), which measures the proportion of role-filler bound to non-null values by a semspec;[10] and (2) *path density* ($\mathrm{density}_p$), an estimate of how densely interconnected a semspec's roles are, *i.e.*, how many average paths (or bindings) there are per role in the semspec.[11] These density measures are combined in the unresolved meaning score $\mathrm{score}\, m_u$:

$$\mathrm{density}_v(ss) = \frac{|\,\mathrm{filled}(ss)|}{|ss|} \tag{4.3.13}$$

$$\mathrm{density}_p(ss) = \frac{|\,\mathrm{bindings}(ss)|}{|ss|} \tag{4.3.14}$$

$$\mathrm{score}_{m_u}(ss) = \frac{\mathrm{density}_v(ss) \cdot f(\mathrm{density}_p(ss))}{2} \tag{4.3.15}$$

where $\mathrm{filled}(ss)$ is the number of non-null roles in ss, $\mathrm{bindings}(ss)$ is the number of identification bindings in ss, and the f function applies a maximum cutoff to the $\mathrm{density}_p$ term and maps it linearly onto the range 0 to 1.

4.3.3 Scoring (resolved) analyses

It is now straightforward to define various scoring functions that can be used by the high-level analysis and resolution processes, based on the various domain-specific evaluation criteria:

- The **unresolved analysis score** $\mathrm{score}(a, u)$ of an analysis a based on utterance u (sans context) is a weighted sum of its scores in the constructional, form and meaning domains:

  $$\mathrm{score}(a, u) = \alpha_c \cdot \mathrm{score}_c(a) \; + \; \alpha_f \cdot \mathrm{score}_f(a, u) \; + \; \alpha_m \cdot \mathrm{score}_{m_u}(ss_a) \quad , \tag{4.3.16}$$

  where α_c, α_f and α_m are parameters chosen to scale the relevant scores.

- The **resolution score** of a resolved semspec $\mathrm{score}(ss, r, z)$ is simply the resolved meaning score $\mathrm{score}_{m_r}(ss, r, z)$:

  $$\mathrm{score}(ss, r, z) = \mathrm{score}_{m_r}(ss, r, z) = \mathrm{fscore}(ss, r, z) \tag{4.3.17}$$

[10] This measure is called *semantic density* in Bryant (2003).

[11] More precisely, the semspec can be partitioned into a set of *binding groups*, where each binding group consists all the roles (denoted as paths from the semspec root level) that are coindexed in the semspec. A higher average number of paths per binding group indicates more internal bindings in the semspec.

- The **resolved analysis score** (or **token score**) $\text{score}(a, r_a, d)$ of an analysis a, its resolution map r_a and input token $d = (u_d, z_d)$ simply substitutes the resolved meaning score in (4.3.17) above into (4.3.16):

$$\text{score}(a, r_a, d) = \alpha_c \cdot \text{score}_c(a) \; + \; \alpha_f \cdot \text{score}_f(a, u_d) \; + \; \alpha_m \cdot \text{score}_{m_r}(ss_a, r_a, z_d) \qquad (4.3.18)$$

- The **fm score** (or **form-meaning score**) $\text{score}_{fm}(a, r_a, d)$ of an analysis a, its resolution map r_a and input token $d = (u_d, z_d)$ is a combined measure of form and meaning coverage, omitting the constructional component from the resolved analysis score.

$$\text{score}_{fm}(a, r_a, d) = \alpha_f \cdot \text{score}_f(a, u_d) \; + \; \alpha_m \cdot \text{score}_{m_r}(ss_a, r_a, z_d) \qquad (4.3.19)$$

Together these criteria provide quantitative measures by which the model's analyzer-resolver evaluates competing analyses and resolution maps, favoring more likely analyses that account for more forms and meanings of a given input token. These measures can be seen as approximating comprehension quality, which are used to evaluate how well a given grammar performs on a token or a corpus of tokens, and to compare how multiple grammars perform on the same corpus. As such, they play a crucial role in the broader context of the learning model, both internal to the model (in guiding the choice among competing constructions) and external to the model (in gauging the learner's progress toward better-performing grammars); these evaluation functions are discussed further in Section 5.2.5.

Example score calculation

The earlier discussion of the input token in Figure 4.4 yielded analysis a = {tiC, tC, bC} (with root construction R_a = THROW-IMP) and resolution map r = {b ⇒ bZ, ds ⇒ dsZ}. The score calculations below are based on the following construction statistics:

$$\text{weight}(\text{Throw-Imp}) = .05 \qquad [\text{Throw-Imp.t} \leftarrow \text{Throw-Cn}] = 10$$
$$[\text{Throw-Imp}] = 10 \qquad [\text{Throw-Imp.y} \leftarrow \text{Ball-Cn}] = 7$$

The form score includes units and relations (weighted equally); a_f accounts for two of the three words and one of three possible binary relations (where only contiguous word order between non-skipped words are taken into account here). The $\text{mapped}_r(ss)$ component of the meaning score includes the two schema instances b and ds, along with the three roles of ds and two identification bindings (*i.e.*, t.throwee $\longleftrightarrow$ b and t.thrower $\longleftrightarrow$ ds.addressee), so $|\,\text{mapped}_r(ss)| = 7$; similarly, $|ss|$ =

9 and $|z|$ = 22. Thus $\text{precision}_m(ss, r, z) = 7/9 = .778$ and $\text{recall}_m(ss, r, z) = 7/22 = .318$. The final score assumes domain weights as follows: α_c = .25, α_f = α_m = 375.

$$\text{score}_c(a) = \text{weight}(\mathbf{r}_a) \times \prod_{\substack{c \in a, \\ n \in \text{constituents}(\mathbf{c})}} \frac{[\mathbf{c}.n \leftarrow \text{type}(c.n)]}{[\mathbf{c}]} = .05 \times \frac{10}{10} \times \frac{7}{10} = .035$$

$$\text{score}_f(a, u) = \frac{|a_f \cap u|}{|u|} = \frac{2+1}{3+3} = .50$$

$$\text{score}_{m_r}(ss, r, z) = \text{fscore}(ss, r, z) = \frac{2 \cdot \text{precision}_m(ss, r, z) \cdot \text{recall}_m(ss, r, z)}{\text{precision}_m(ss, r, z) + \text{recall}_m(ss, r, z)} = \frac{2 \times .778 \times .318}{.778 + .318} = .451$$

$$\begin{aligned} \text{score}(a, r, d) &= \alpha_C \cdot \text{score}_c(a) + \alpha_f \cdot \text{score}_f(a, u) + \alpha_m \cdot \text{score}_{m_r}(ss_a, r, z) \\ &= \alpha_c \times .035 + \alpha_f \times .50 + \alpha_m \times .432 \\ &= .25 \times .035 + .375 \times .50 + .375 \times .432 = .358 \end{aligned}$$

It is instructive to consider the score of a competing analysis not including the THROW-IMP construction, *i.e.*, where a' = {tC, bC} and r' = {b ⇒ bZ}. Since there is no single root, the constructional score makes use of the provisional root $\mathbf{r}_{a'}$. Assuming the total count across the constructions of the mini-grammar is 100, $\text{weight}(\mathbf{r}_{a'}) = .01$. The form and meaning domains of a' are relatively sparser than a. The same words are covered, but no order is specified between them. The Throw and Ball meanings are present, but without any bindings between them, and no DiscourseSpace schema is explicitly evoked. Thus $|\text{mapped}_r(ss)| = 1$ (just b) and $|ss| = 4$ (with an unchanged $|z| = 22$), yielding $\text{precision}_m(ss', r', z) = 1/4 = .25$ and $\text{recall}_m(ss', r', z) = 1/22 = .045$. Recalculating:

$$\text{score}_c(a') = .01 \times 1 \times 1 = .01$$

$$\text{score}_f(a', u) = \frac{2+0}{3+3} = .333$$

$$\text{score}_{m_{r'}}(ss', r', z) = \frac{2 \times .25 \times .045}{.25 + .045} = .076$$

$$\begin{aligned} \text{score}(a', r', d) &= \alpha_c \times .01 + \alpha_f \times .333 + \alpha_m \times .076 \\ &= .25 \times .01 + .375 \times .333 + .375 \times .076 = .156 \end{aligned}$$

As expected, this lexical-only analysis scores worse in all domains: its constructional graph is less likely, and its coverage of the relevant forms and meanings is much less complete.

Structural and referential ambiguity resolution

The input token in Figure 4.10 is similar to that in Figure 4.4, pairing the utterance *throw the ball over here* with a context including two specific balls, one located near (bZ) and the other far (b2Z) from the mother. The utterance is a simplified version of the structurally ambiguous example (4–2) (*Get the green one over there*) discussed earlier, with two possible interpretations:

$$\left(\begin{bmatrix}\text{Utterance} & \\ \text{text}: & \text{“throw the ball over here”} \\ \text{intonation}: & \text{falling}\end{bmatrix},\ \begin{bmatrix}\text{Context} & \\ \text{participants}: & \text{Mother, Naomi, bZ:Ball, b2Z:Ball} \\ \text{scene}: & \begin{bmatrix}\text{tlZ:Trajector-Landmark} & \\ \text{trajector}: & \text{bZ:Ball} \\ \text{landmark}: & \text{Mother} \\ \text{relation}: & \text{proximal}\end{bmatrix}, \begin{bmatrix}\text{tl2Z:Trajector-Landmark} & \\ \text{trajector}: & \text{b2Z:Ball} \\ \text{landmark}: & \text{Mother} \\ \text{relation}: & \text{distal}\end{bmatrix} \\ \text{discourse}: & \begin{bmatrix}\text{dsZ:DiscourseSpace} & \\ \text{speaker}: & \text{Mother} \\ \text{addressee}: & \text{Naomi} \\ \text{speech act}: & \text{imperative} \\ \text{joint attention}: & \text{bZ:Ball} \\ \text{activity}: & \text{play} \\ \text{temporality}: & \text{ongoing}\end{bmatrix}\end{bmatrix}\right)$$

Figure 4.10. A structurally and referentially ambiguous input token based on the utterance “throw the ball over here”.

(4–3) Throw the ball over here.

a. [Throw [the ball [over here]]] . *(transitive: proximal referent, unspecified goal)*
(≈ “(You) throw the ball that is here.”)

b. [Throw [the ball] [over here]] . *(caused motion: proximal goal)*
(≈ “(You) throw the ball and cause it to move here.”)

The structural ambiguity above concerns the phrase *over here*, and whether it is a locational modifier within the referring expression *the ball over here* or a goal specifier for a caused motion construction. But this ambiguity arises only when the requisite constructions are present in the grammar. Below we consider analyses using first a grammar that lacks these constructions and then a grammar that has them.

Referential ambiguity. Figure 4.11 shows the result of analyzing the token using the mini-grammar G_1 from the previous example, supplemented with a HERE-CN construction that represents “proximal to the speaker” using a Trajector-Landmark schema whose landmark role is bound to the evoked discourse speaker. The chart yields a (single) partial analysis with two roots (and two skipped forms). But contextual resolution produces two candidate maps, since either of the two balls in context could be the referent of b. It is thus referentially ambiguous despite being structurally unambiguous.

Structural ambiguity. Assume a grammar G_2 that contains the structures of G_1, the HERE-CN defined above, the DETERMINEDNOUN-EXPR construction from Figure 3.16, and the following:[12]

- a lexically specific OVERHERE-CN construction that links the forms *over here* with a deictically constrained location, *i.e.*, a location near the speaker;

[12] ECG definitions for these constructions are left as an exercise for the interested reader.

ti1C: Throw-Imp [tC, bC; sk=1]				
tC: Throw-Cn		**bC: Ball-Cn**		**hC: Here-Cn**
`throw`	`the`	`ball`	`over`	`here`

$$\begin{bmatrix} \text{Semspec} & & \\ \text{participants}: & \text{b:Ball} & \\ \text{scene}: & \begin{bmatrix} \text{t:Throw} & \\ \text{thrower}: & \text{ad} \\ \text{throwee}: & \text{b} \end{bmatrix}, & \begin{bmatrix} \text{tl:Trajector-Landmark} & \\ \text{trajector}: & \text{tr:Entity} \\ \text{landmark}: & \text{sp} \\ \text{relation}: & \text{proximal} \end{bmatrix} \\ \text{discourse}: & \begin{bmatrix} \text{ds:DiscourseSpace} & \\ \text{speaker}: & \text{sp:Human} \\ \text{addressee}: & \text{ad:Human} \\ \text{speech act}: & \text{imperative} \end{bmatrix} & \end{bmatrix}$$

Analysis A1: Partial [tiC, hC; sk = 2]
Resolution map R1-1: $\{$ b $\Rightarrow$ bZ, ds $\Rightarrow$ dsZ, tl $\Rightarrow$ tlZ $\}$
Resolution map R1-2: $\{$ b $\Rightarrow$ b2Z, ds $\Rightarrow$ dsZ, tl $\Rightarrow$ tlZ$\}$

Figure 4.11. Unambiguous analysis with ambiguous resolution: results of analyzing *throw the ball over here*, yielding a single partial analysis and two candidate resolution maps.

- a (partially) lexically specific DETERMINEDNOUNOVERHERE-EXPR construction that resembles the DETERMINEDNOUN-EXPR construction, but extends it with an additional OVERHERE-CN constituent that constrains the location of the determined noun; and
- a clause-level CAUSEDMOTION-IMP construction similar to the lexically specific THROW-IMP but appropriate for expressing a causing action, theme object and goal location.

The results of analyzing the token with G_2 are shown in Figure 4.12, where constructs included in either of the two spanning analyses with a single root are displayed in bold. (Multiple instances of both clausal constructions are shown in the chart; these include instances that skip one or more forms (such as the determiner *the* or the preposition *over*) or span only part of the utterance.) The semspecs for the two single-root analyses (not pictured) are very similar to the semspec shown in Figure 4.11, differing only in the binding of the Trajector-Landmark schema's trajector role: the transitive analysis A1 binds this to b (which is also bound to the ball to be thrown), while the caused-motion analysis A2 binds this to loc (the goal location of the throwing event).

The figure also shows the candidate resolution maps available for each analysis. Unlike the analysis using G_1, the semspecs for both A1 and A2 here have additional constraints that limit the number of candidate resolution maps. In the case of A1, the binding between the thrown ball and the proximal ball rules out b2Z as a possible referent of b, resulting in only a single resolution map with three bindings. In the case of A2, both referents remain possible, just as in the G_1 analysis. Now, however, the specification of the throwing event's goal being proximal to the speaker results

<table>
<tr><td colspan="5">cmi3C: CausedMotion-Imp [tC, dnC, ohC]</td></tr>
<tr><td colspan="5">cmi2C: CausedMotion-Imp [tC, dnC, hC; sk=1]</td></tr>
<tr><td colspan="5">cmi1C: CausedMotion-Imp [tC, bC, hC; sk=2]</td></tr>
<tr><td colspan="5">ti3C: Throw-Imp [tC, dnohC]</td></tr>
<tr><td colspan="3">ti2C: Throw-Imp [tC, dnC]</td><td></td><td></td></tr>
<tr><td colspan="3">ti1C: Throw-Imp [tC, bC; sk=1]</td><td></td><td></td></tr>
<tr><td></td><td colspan="4">dnohC: DeterminedNounOverHere-Expr [thC, bC, ohC]</td></tr>
<tr><td></td><td colspan="2">dnC: DeterminedNoun-Expr [thC, bC]</td><td colspan="2">ohC: OverHere-Cn [oC, hC]</td></tr>
<tr><td>tC: Throw-Cn</td><td>thC: The</td><td>bC: Ball-Cn</td><td>oC: Over-Cn</td><td>hC: Here-Cn</td></tr>
<tr><td>throw</td><td>the</td><td>ball</td><td>over</td><td>here</td></tr>
</table>

Analysis A1: ti3C [tC, dnohC [thC, bC, ohC [oC, hC]]]
Resolution map R1-1: { b ⇒ bZ, ds ⇒ dsZ, tl ⇒ tlZ }

Analysis A2: cmi3C [tC, dnC [thC, bC], ohC [oC, hC]]
Resolution map R2-1: { b ⇒ bZ, ds ⇒ dsZ }
Resolution map R2-2: { b ⇒ b2Z, ds ⇒ dsZ }

Figure 4.12. Ambiguous analysis: chart resulting from analysis of *throw the ball over here*. Input grammar includes constructions for determined nouns (*the ball*, *the ball over there*) and an imperative caused motion predicates.

in no contextually accessible referents for tl. There are thus two resolution maps possible for A2, each with two bindings.[13]

Which of these analyses is chosen depends on a number of factors, including constructional usage statistics and the relative weights applied to each domain. The two analyses cover the same forms, but A1's resolution map is larger than either of A2's resolution maps and thus has a higher meaning score. Assuming that the caused-motion analysis (A2) has a higher root frequency and construction score than the transitive analysis (A1), the winner here depends on whether having an edge in the meaning domain or the constructional domain counts more. For current purposes, the example demonstrates how the integrated analysis and resolution processes can successfully handle structural and referential ambiguity, with constraints from both helping to find the minimally ambiguous overall solution.

* * *

The goal of this chapter has been to render explicit several key components of the language understanding model depicted in Figure 4.1. Each of the main structures involved in the two processes of constructional analysis and contextual resolution has been defined, including the formal

[13]One of these resolution maps is certainly better than the other, based on the current locations of the respective balls and the specification that the goal location is near the speaker. That is, the interpretation corresponding to R2-1 specifies that the addressee should throw ball bZ from its current position to a goal that is also near the speaker, which is inferentially coherent than R2-2's requested throwing of b2Z from a more distant source location to the same goal. Such inference-based resolution ranking may depend on interactions with the simulation engine and detailed event knowledge, which falls outside of the scope of the simple resolution process described here.

ECG structures introduced in Chapter 3; the input utterance and context; the analysis and semspec; and the resolution map and resolved semspec. The simple implementations of these processes presented here satisfy the basic input-output requirements specified by these structures. Moreover, they are designed to produce robust partial analyses using incomplete grammars and to resolve structural and referential ambiguity using both specifically linguistic and situational constraints.

While the particular analyzer and resolution algorithms offered here do not address all the complexities involved in language comprehension in general, they do demonstrate how an embodied construction grammar can be used to interpret and resolve utterances. They thus serve as a rudimentary theory of language understanding that is adequate for a subset of language comprehension, appropriate for child language and designed to accompany the theory of language structure proposed in Chapter 3. Most importantly for current purposes, the structures yielded by these processes and the evaluation metrics by which they are ranked highlight precisely how well and in what respects a given grammar accounts for input data, thus providing a concrete basis for proposing and evaluating candidate analyses — and, by extension, candidate constructions and grammars. Both of these functions — hypothesizing new structures and choosing among them — are exploited in the usage-based theory of language acquisition to follow.

Chapter 5

Learning embodied constructions

It isn't that they can't see the solution. It's that they can't see the problem.
— G.K. Chesterton

We are now in a position to address ourselves more directly to the learning task introduced in Chapter 1. Section 5.1 recapitulates the most relevant constraints encountered in the foregoing chapters. These are distilled in Section 5.2 into a formal statement of a class of grammar learning problems, along with a specific instance of this class. The focus then shifts to seeking an adequate solution: Section 5.3 reviews some candidate approaches, and Section 5.4 lays out a framework that adapts these to satisfy our problem constraints.

5.1 The child's learning task: a review

The developmental and linguistic evidence reviewed thus far suggests the following informal synopsis of the language learning task: Children are situated in rich experiential contexts, subject to the flow of unfolding events. At all stages, they exploit a panoply of cognitive and social abilities to make sense of these experiences. To make sense of *linguistic* events — sounds and gestures used in their environments for communicative purposes — they also draw on mappings between these perceived forms and conceptual knowledge. These mappings, or **constructions**, are typically incomplete with respect to the utterances they encounter, but even an imperfect understanding of an utterance provides direct clues to its intended meaning and thereby reduces the burden on pragmatic inference and situational context. The goal of language learning, from this perspective, is to acquire an increasingly *useful* set of constructions (or **grammar**) — that is, one that allows accurate interpretations of utterances in context with minimal recourse to general inference procedures. In the limit, the grammar should stabilize, while facilitating the comprehension of both familiar and novel input.

This section summarizes the most relevant constraints on the problem. These fall into three categories: structural constraints arising from the nature of the target of learning; usage-based constraints arising from the goal of facilitating language comprehension; and cognitive and developmental constraints arising from the nature of the learner.

5.1.1 Representational requirements

The target of learning is a construction-based grammar in the general sense described in Section 2.2.1, as instantiated by the Embodied Construction Grammar formalism.

- Constructions are *cross-domain*: they are mappings over the domains of form and meaning. Units of form and meaning are *schematic* descriptions summarizing the linguistically relevant aspects of more detailed embodied representations (such as acoustic or auditory representations for form, and motor, perceptual and other conceptual representations for meaning).
- Constructions may have *constituents*, *i.e.*, subcomponents that are themselves constructions.
- Constructions are themselves complex *categories*, organized in a *typed multiple inheritance hierarchy*. They may be subcases of other constructions, inheriting structures and constraints. Their constituent constructions, as well as their form and meaning poles, may also be typed.[1]
- Constructions may include *relational constraints*, i.e. relations that must hold of form and/or meaning elements (such as binding constraints and ordering constraints).

The focus here is on constructions licensing multiword expressions, which typically involve multiple forms, meanings and constituents, along with relational constraints among these. The problem thus demands an approach to learning that can accommodate the representational complexity imposed by structured cross-domain mappings.

[1] As noted in Section 3.1.4, the choice of an inheritance-based type hierarchy, while a reasonable simplification for current purposes, does not capture the graded, radial nature of constructional organization.

5.1.2 Comprehension-based constraints

The framework described in Chapter 4 distinguishes several processes involved in language comprehension:

- *constructional analysis*: identifying the set of constructions instantiated by a given utterance, and its corresponding *semantic specification* (or *semspec*);
- *contextual resolution*: mapping objects and events in the semspec to items in the current communicative context, producing a *resolution map* and a *resolved semspec*; and
- *simulative inference*: invoking the dynamic embodied structures specified by the (resolved) semspec to yield contextually appropriate inferences.

As discussed in Chapter 4, these processes must tolerate uncertainty, ambiguity and noise: there may be multiple possible constructional analyses for a given utterance and multiple ways of resolving referents to the context; there may be errors in the perceived input utterance or communicative context; and some inferences may be only probabilistically licensed. Comprehension processes must also cope gracefully with input not covered by its current grammar: all constructional analyses, including those that account for the entire input utterance and those that do not (*i.e.*, *partial* analyses), should produce (partial) semspecs that can be resolved and simulated.

These considerations suggest that comprehension is not a binary matter, but rather one of degree: interpretations can be judged as relatively more or less complete, coherent and consistent with the context. That is, the language comprehension processes above require some means of evaluating candidate interpretations and choosing those that contribute to effective and efficient comprehension (*e.g.*, by maximizing utterance interpretability in context, or minimizing constructional and contextual ambiguity). Since progress in learning can be judged only by improvement in comprehension ability, the same facilitating factors and evaluation criteria should also serve to guide and constrain the learning process.

5.1.3 Developmental desiderata

As reviewed in Section 2.1.3, children entering the multiword stage have amassed a battery of sensorimotor, conceptual, pragmatic and communicative skills, including:

- familiarity with a variety of people, objects, locations, and actions, including both specific known entities (*e.g.*, mother, favorite stuffed toy, bed) and more general categories of objects and actions (*e.g.*, milk, blocks, grasping, being picked up);

- ability to infer referential and communicative intent, including determining objects of (joint) attention and speech act type;
- familiarity with intonational contours and their associated speech acts;
- relatively well-developed ability to segment utterances into words; and
- a growing inventory of lexical mappings, including labels for many familiar ontological categories, as well as social and relational words.

The timelines for these ongoing developmental and learning processes overlap with the acquisition of multiword constructions, and they vary dramatically across individual children. But all of these diverse kinds of information may, in principle, be available by the time children begin to learn their first word combinations.

Several trends in the acquisition of multiword constructions have also been identified, as reviewed in Section 2.2.3. Typically, children:

- learn more specific constructions before more general constructions, more frequent constructions before less frequent constructions, and smaller constructions before larger constructions;
- require a relatively modest amount of data to achieve a baseline level of performance;
- receive relatively little negative evidence in the form of error correction; and
- generalize constructions to arguments beyond those observed in the input, both appropriately and inappropriately.

Again, these patterns are subject to variation, both across and within individuals. Some children learn larger multiword chunks as a single unit before later reanalyzing them in terms of their constituent pieces; some persist in producing one-word utterances despite being able to string several of these together (separated prosodically) to express complex predications.

These developmental findings inform the kinds of behaviors that should be exhibited by a successful model of construction learning. That is, the course of acquisition should be qualitatively comparable to that of a child at a similar developmental stage, and the model should be flexible enough to encompass multiple learning styles.[2]

[2] I elide the distinction between comprehension and production here; while the timeline of acquisition may differ between these, I assume that many aspects of the course of acquisition are roughly comparable.

5.2 Problem statement(s)

This section translates the constraints just reviewed into a formal statement of a class of grammar learning problems. Section 5.2.1 identifies the main structures and processes involved (defined in detail in Section 4.1) and uses them to define the general problem of comprehension-based language learning; these are then elaborated with respect to the problem of learning relational constructions in Sections 5.2.2-5.2.5, culminating in a more specific problem statement in Section 5.2.6.

5.2.1 Construction learning defined

The primary representational structures involved in language learning are as follows:

- A **grammar** G is a pair (S, C), where S is a **schema set**, and C is a **construction set** defined with respect to S. Both S and C are typed multiple inheritance hierarchies whose members are represented using the ECG schema and construction formalisms, respectively.
- Schemas in S provide parameters to embodied structures in a separate **ontology** O of detailed sensorimotor representations and other structures that are not specifically linguistic.[3]
- A **corpus** D is an ordered list of **input token**s d. Each input token d is a pair (u, z), where u is an **utterance** and z is a communicative **context**.

Language comprehension is divided into several interrelated processes:

- A **construction analyzer** provides the function $\mathrm{analyze}(d, G)$, which takes an input token $d = (u, z)$ and a grammar G and produces an **analysis** a_d and an associated **semspec** $ss_a = \mathrm{semspec}(a_d)$.
- A **contextual resolver** provides the function $\mathrm{resolve}(ss, z)$,which takes a semspec ss and a context z and produces a **resolution map** r and a **resolved semspec** rss.
- A **simulation engine** provides the function $\mathrm{simulate}(ss, z)$, which takes a resolved semspec rss as input parameters for a dynamic simulation with respect to the context z, using structures in the conceptual ontology O to produce a set $\mathrm{inferences}(rss)$ of simulation inferences.

These processes evaluate their output structures using the folllowing quantitative scoring criteria:[4]

[3]The sense of 'ontology' intended here is broader than that typically used to represent static taxonomic relations.
[4]Probabilistically minded readers may note a suggestive resemblance between the scoring functions listed here and the Bayesian prior probability of a grammar, likelihood of observed data with respect to a grammar, and posterior probability of a grammar given the data, respectively). This family resemblance is further elucidated in Chapter 7.

- $\mathrm{score}(G)$ is the **grammar score**, which measures intrinsic properties of a grammar G. Generally, it is defined to reward simpler, more compact grammars.

- $\mathrm{score}(d|G)$ is the **token score**, which measures how well token d is comprehended using grammar G. Generally, it is defined to reward simpler, more likely analyses that account for more of the input forms and meanings and are more easily interpreted in context.

- $\mathrm{score}(D|G)$ is the **corpus score**, which aggregates token scores $\mathrm{score}(d|G)$ based on grammar G over the entire corpus D.

- $\mathrm{score}(G|D)$ is the **grammar performance score**, which measures how well a grammar G accounts for a corpus D. This term should incorporate aspects of both the (intrinsic) grammar score and its associated corpus score.

The terms above allow us to characterize a general class of language learning problems consistent with the goals and constraints established in the preceding chapters.

Comprehension-based language learning

hypothesis space The target of learning is a grammar G defined using the ECG formalism.

prior knowledge The learner has an initial grammar $G_0 = (S_0, C_0)$, a constructional analyzer that performs the $\mathrm{analyze}$ and $\mathrm{resolve}$ functions, a simulation engine that performs the $\mathrm{simulate}$ function, and a conceptual ontology O.

input data The learner encounters a sequence of input tokens from a training corpus D.

performance criteria The grammar performance score of successive grammars G_i should improve and eventually stabilize as more data is encountered.

The language learning problem as defined here admits many instantiations, corresponding to different ways in which the structures and processes involved can be further delineated. A version of the problem focusing on lexical learning, for example, may restrict the initial and target grammars to lexical constructions; the precise form of the input data (raw sensory data, phonological or phonemic sequences, orthographic strings) may be chosen to reflect differing assumptions about the perceptual abilities of the learner or the starting point of learning; and the scoring functions may vary in the aspects of the grammar and analysis they reward.

My focus in this work is on the comprehension-driven acquisition of relational constructions — *i.e.*, the emergence of constituent structure. I thus restrict each component of the broader language learning problem with some simplifying assumptions, as described in the next several sections.

5.2.2 Hypothesis space

The hypothesis space for language learning is defined as a grammar $G = (S, C)$, since both S and C are taken to be part of linguistic knowledge. Recall that the schema set S contains the basic units of linguistic form and meaning, as discussed in Section 3.2. These are the two domains linked by all linguistic constructions, which together encompass all *linguistically* relevant concepts, including both lexicalizable and grammaticizable notions in the sense discussed in Section 2.2.2 (Talmy 2000; Slobin 1997).

I distinguish such concepts from the set O of *all* ontologically available categories, concepts and domain-based knowledge. O is taken to encompass the full range of human conceptual and perceptual potential, from multimodal, biologically grounded structures used in action, perception, emotion and other aspects of experience to the more abstract but often richly structured categories and relations in less directly embodied domains. The schemas in S summarize (a subset of) the structures of O, and they exhibit noticeable regularities across languages (as might be expected based on their putative common embodied basis). In the current work, however, I assume that the actual set of linguistically relevant features, as well as the constellations of features that are regularly packaged for particular linguistic purposes, must be learned on a language-specific basis.

Different learning processes are associated with each of these structures:

- **concept learning**: the formation of *ontological categories* of O based on embodied experience;
- **(linguistic) schema learning**: the reification of ontological categories and features of O into the *linguistically relevant categories* of S, *i.e.*, a set of embodied schemas for the domains of form and meaning;
- **construction learning**: the pairing of form and meaning schemas of S into the *specifically linguistic categories* of C.

These processes occur in parallel through development and into adulthood, and the relationships among them are complex. There is a natural dependency of schema learning on concept learning (a schema being a specific kind of concept), and of construction learning on schema learning (a construction being a pairing of form and meaning schemas). But structures arising in each kind of learning may also exert a mutually reinforcing influence on those arising in the others. A strong Whorfian position, for example, might assert a determinative effect of S on O; a weaker stance might posit a more balanced relationship between them. Likewise, it may be precisely the regularity of certain form-meaning pairs in C that elevates their associated concepts into the schemas of S.

The current work does not attempt to model all of these concurrent interactions; rather, these processes are idealized as independent but interleaved. Specifically, I assume that construction learning can be usefully modeled as taking place with respect to a *fixed* schema set and conceptual ontology. The space of possible grammars G thus reduces to the space of possible construction sets C defined with respect to S, with particular emphasis on the acquisition of structured mappings.

Note that this formulation nonetheless presumes significant *indirect* interactions based on the inherent structural connections among concepts, schemas and constructions: constructions refer directly to form and meaning schemas, and indirectly (via schemas) to embodied structures and general ontological categories. Any changes to these structures (due to interleaved concept and schema learning) could potentially affect the course of construction learning, where shorter interleaving intervals reflect a tighter connection. In the limit, this interleaving could theoretically be on a per-token basis. Potential extensions to the model to permit more direct interactions among these processes are addressed in Chapter 9.

5.2.3 Prior knowledge

The learner is assumed to have access to both a grammar (with schemas and constructions) and a language understanding model (performing constructional analysis and contextual resolution); these correspond to naive theories of language structure and use appropriate for supporting language behaviors observed during the single-word stage. No claims are made here about whether or to what extent these structures and processes are innate, or simply acquired before and during single-word learning.

Schemas and constructions

Schemas and constructions are represented using the ECG formalism, as described in Chapter 3 and Section 4.1.1. Assumptions about the contents of the initial grammar depend on the particular phenomena or learning stage of interest; the specifications below reflect the current focus on the earliest relational constructions in English:

- Word forms are represented as orthographic strings. Intonational information is represented using an intonation role filled by one of a limited set of familiar contours (falling, rising and neutral). Potential form relations are limited to the ordering relations before and meets.

- Meaning schemas reflect a typical child's ontological knowledge of concrete entities, actions, scenes and relations (such as those shown in Figure 3.8).

- The initial construction set consists of simple mappings linking word forms to embodied schemas. These include concrete referring expressions (for people, objects and places), action words and spatial relations (like the ones defined in Figure 4.3). No other function words (*e.g.*, determiners), relational constructions or other explicitly grammatical categories are present at the outset of learning.

- The initial construction set has a relatively flat organization. For example, no type distinction is initially made between common nouns and proper nouns, though this distinction could emerge through learning.

The inclusion of (some) simple lexical mappings in the initial grammar is not intended to suggest that all lexical learning precedes relational learning; in fact, the model assumes that both kinds of constructions can be learned based on the same general operations (see Section 6.2). But our current focus is on the particular challenges of learning early relational constructions; by this stage many lexical constructions will have been acquired by any of various learning strategies that have been proposed. This claim may be more reasonable for some conceptual domains (*e.g.*, object labels and names) than others (*e.g.*, verbs and spatial relations terms); but even inherently relational terms must be learned in part by association strategies; see Chapter 9 for further discussion of how and whether these assumptions can be relaxed.

Processes of language comprehension

The inclusion of language comprehension as part of the learner's prerequisite abilities follows naturally from the usage-based assumptions of our broader scientific framework. The simulation-based framework set forth in Chapter 3 proposes that language understanding is driven by active, context-sensitive mental simulations. A fully integrated model of (adult) language comprehension, including analysis, resolution and simulation, remains the subject of ongoing research. For the language learning stage under investigation, the simplified model of language understanding described in Chapter 4 is sufficient:

- The constructional analyzer takes input tokens (as described in Section 5.2.4) and an ECG grammar and produces constructional analyses and their corresponding semspecs. Analyses are ranked according to the scoring criteria that evaluate each of the constructional, form and meaning domains.

- The contextual resolver finds the best resolution map between a semspec and a given input context. No context history across utterances is assumed, and no explicit creation of referents (*e.g.*, based on indefinite referring expressions) is currently allowed. Resolution maps are ranked according to scoring criteria that rewards maps that have the greatest alignment between the semspec and context.

These language understanding processes satisfy the key requirements of the learning model: they must perform robustly with incomplete grammars; they must provide partial analyses and resolution maps that motivate the formation of new constructions; and they must provide a means of ranking and evaluating analyses.

5.2.4 Input data

The input to learning is characterized as a corpus of **input tokens**, each consisting of an utterance paired with its communicative context (as described in Section 4.1.2). The contents of both utterance and context are intended to reflect the sensorimotor, conceptual and inferential abilities of a human learner at the relevant stage, as reviewed in Chapter 2. The most relevant assumptions are as follows:

- The learner can pragmatically filter the glut of continuous percepts in the environment to extract those objects, relations and intentions most relevant to the attended scene and chunk them into discrete participants and events (*i.e.*, perform scene parsing).
- Context items instantiate schemas in the schema set S. The choice of S (as opposed to the ontology O) as the basis for contextual representation is consistent with the current focus on construction learning, as opposed to (linguistic) schema learning. The assumption is that linguistically relevant features may be more salient than others for the purposes of understanding and learning language.[5]
- The learner can reliably associate utterances with the appropriate scenes or referents, irrespective of their precise temporal alignment (*i.e.*, whether the utterance occurred before, during or after the event). In some cases, as in the imperative speech act of the example situation, the associated event may be inferred from context, whether or not the desired event ultimately takes place.
- The learner receives only positive examples, in the sense that tokens are not identified as non-occurring.

[5] This claim might be considered a comprehension-based analogue to Slobin's (1991) "thinking for speaking" hypothesis.

- No extended contextual history is represented in the model. A limited notion of shared situational context is available: each input token is associated with an *episode*, which has a persistent set of (shared) episode participants. In general, however, input tokens are treated by the analysis, resolution and learning processes as standalone units with direct access to any relevant contextual information.

Figure 5.1 shows an example input token similar to the earlier one in Figure 4.4. The same utterance "throw the ball" (with falling intonation) is paired with a context that explicitly represents the (desired) throwing schema requested of the child (Naomi) by the mother. The inclusion of a Throw context item reflects the assumptions above: namely, that the child can infer the mother's communicative intention from pragmatic cues (*e.g.*, gaze, gesture).

$$\left(\left[\begin{array}{ll}\textsf{Utterance} & \\ \textsf{text :} & \textsf{“throw the ball”} \\ \textsf{intonation :} & \textsf{falling}\end{array}\right],\ \left[\begin{array}{ll}\textsf{Context} & \\ \textsf{participants :} & \textsf{Mother, Naomi, Ball} \\ \textsf{scene :} & \left[\begin{array}{ll}\textsf{Throw} & \\ \textsf{thrower :} & \textsf{Naomi} \\ \textsf{throwee :} & \textsf{Ball}\end{array}\right] \\ \textsf{discourse :} & \left[\begin{array}{ll}\textsf{DiscourseSpace} & \\ \textsf{speaker :} & \textsf{Mother} \\ \textsf{addressee :} & \textsf{Naomi} \\ \textsf{speech act :} & \textsf{imperative} \\ \textsf{joint attention :} & \textsf{Ball} \\ \textsf{activity :} & \textsf{play} \\ \textsf{temporality :} & \textsf{ongoing}\end{array}\right]\end{array}\right]\right)$$

Figure 5.1. A typical input token for learning: the utterance "throw the ball" paired with a communicative context in which the Mother tells the child to throw the ball.

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), annotated by developmental psychologists as part of a study of motion utterances (May *et al.* 1996). These annotations indicate semantic and pragmatic features available in the scene; they are described in more detail in Section 8.1.1.

5.2.5 Performance criteria

Under the assumption that communicative competence is a goal (if not necessarily the only goal) of language learning, the evaluation criteria used for measuring comprehension can also be exploited to gauge the overall progress of the learner. By testing the learning model at regular intervals during training, we can assess how new constructions incrementally improve its ability to comprehend new input tokens.

In general, any measure associated with a more broadly encompassing theory of language use could be incorporated into the grammar performance score, rewarding communicative success in both comprehension and production, or perhaps integrating an explicit notion of agent utility or goal achievement. For example, given a model of comprehension with access to a simulation engine, the learner's performance score could include a measure of the (contextually appropriate) inferences resulting from simulation. Likewise, a model capable of language production could include a measure of how well a produced utterance expresses a given communicative intention, or whether it succeeds in achieving a particular agent goal.

The performance criteria used for the current model are based on those defined in Section 4.3 for the analysis and resolution processes. Recall that those measures allow the analyzer and resolver to rank candidate analyses and resolution maps. When taken in aggregate (or averaged) across the input tokens of a test corpus, however, they also provide a measure that can be compared across grammars. The most relevant criteria are the following:

- *Token score:* The *token score* (*i.e.*, the resolved analysis score) serves as an overall indication of comprehension, since it aggregates constructional, form and meaning components resulting from both analysis and resolution.

- *Form-meaning coverage:* The *form-meaning score*, along with the domain-specific precision and recall scores on which it is based, provides a measure of how completely a grammar accounts for input utterances and contexts.

- *Analysis size:* The absolute size of analyses (*i.e.*, the number of form units and relations $|a_f|$ and the number of meaning schemas and bindings $|a_m| = |ss|$ accounted for) may increase as the grammars improve. Better grammars should also produce more complete analyses (*i.e.*, with a single spanning root construct) and a lower average number of roots.

The measures above provide a quantitative basis for evaluating grammar performance. As discussed further in Chapter 8, it is also useful to make qualitative assessments of the kinds of constructions learned and errors produced by the model, relative to observed patterns of child language acquisition.

5.2.6 Summary: relational construction learning

The specific problem faced by the learner in acquiring relational constructions can be summarized as follows:

Comprehension-based learning of relational constructions

hypothesis space The target of learning is a grammar $G = (S, C)$ defined using the ECG formalism, where S is a fixed schema set.

prior knowledge The learner has an initial grammar $G_0 = (S_0, C_0)$, with C_0 consisting of lexical constructions (for concrete objects, actions and relations) and a language understanding model that performs constructional analysis and contextual resolution.

input data The learner encounters a sequence of input tokens from a training corpus D, each represented as a pair of discrete feature structures describing the utterance form and its communicative context.

performance criteria The grammar performance of successive grammars G_i on a test corpus is measured by improvement in comprehension, using the criteria from the language understanding model including form-meaning coverage, analysis size, completeness and ambiguity.

This view of the core computational problem departs significantly from traditional framings of grammatical induction. Each aspect of the problem fundamentally incorporates meaning and use, and the choice of a construction-based grammar as the target of learning introduces structural complexity in each of the constructional, form and meaning domains. But the challenges of adopting such rich representations are more than offset by the advantages of including similarly rich structures corresponding to the learner's prior conceptual, linguistic and pragmatic knowledge. Moreover, the learner has access to comprehension processes with which to make sense of input utterances in context.

The remainder of this chapter explores solutions that exploit both of these factors — the richer representations available to the learner, and the language comprehension process itself — to overcome the greater inherent representational complexity of learning relational constructions.

5.3 Approaches to learning

The problem as set forth above presents a combination of challenges not directly addressed by previous approaches:

- The target of learning is an inventory of constructions on a continuum of abstraction and size, including both lexically specific and more general patterns, and both simple and structured mappings across the form and meaning domains.
- The space of possibilities is not pre-specified in terms of a fixed set of parameters.
- The goal of learning is to move toward grammars that allow progressively better comprehension of new input.

- The learning strategy must be incremental and sensitive to statistical regularities in the data.

Together these constraints preclude an exhaustive search through all possible grammars for a uniquely specified "correct" grammar. Rather, the topology of the space of grammars must be discovered (or constructed) on the basis of experience. But a purely bottom-up, instance-based approach will also not suffice, since some ability to generalize beyond seen data is required. We thus seek a learning algorithm that strategically converges on simpler (and more general) grammars without unduly sacrificing closeness of fit to observed data.

As suggested in Section 2.3.4, machine learning techniques in the closely related Bayesian and information-theoretic traditions provide the most natural candidates for capturing this tradeoff. The approach taken here can be seen as adapting previous work along these lines to accommodate the structural complexity of the target ECG representation and exploit the tight integration between language learning and understanding. After reviewing the basic Bayesian approach to model selection (Section 5.3.1), this section describes key features of the most relevant previous model of language learning, Bayesian model merging (Section 5.3.2) and then identifies the challenges involved in adapting these to the current task (Section 5.3.3).

5.3.1 Bayesian inference

Bayesian methods are by this point well-established both within the AI community and in applications that range across a swath of scientific realms. They are amply documented in both the literature (Russell & Norvig 1994; Jurafsky & Martin 2000); here I briefly review the probabilistic basis for Bayesian inference. In general, probability theory can be applied to assess the probability of a given event, and by extension the relative probabilities of various events; the event in question may also be a candidate hypothesis about some phenomenon.

In particular, the probabilistic event may be a hypothesis about model structure, where *model* refers in a general sense to any account for some observed data. That is, the goal is to select the most likely model M given the data D, *i.e.*, the one with the highest *conditional* probability $P(M|D)$. Though it is in principle possible to estimate $P(M|D)$ directly, it is sometimes more convenient to exploit the relation between the joint and conditional probability, expressed in (5.3.1), to produce the alternate expression for $P(M|D)$ in (5.3.2):

$$P(M \cap D) = P(M) \cdot P(D|M) = P(D) \cdot P(M|D) \tag{5.3.1}$$

$$P(M|D) = \frac{P(M) \cdot P(D|M)}{P(D)} \tag{5.3.2}$$

The Bayesian insight is that this rearrangement of the terms allows one to estimate $P(M|D)$ (which in this formulation is called the *posterior* probability) in terms of other quantities that are often more convenient to obtain, namely: the *prior* expectation $P(M)$ over what the model should be; the *likelihood* $P(D|M)$ that a given model M would give rise to the actual data D encountered; and a prior expectation $P(D)$ over the data itself.

The prior $P(M)$ is typically defined to encode biases dependent on the model space, often favoring simpler models over more complex models, and $P(D|M)$ is typically directly provided by the model. In choosing the best model $\hat{M}$ among all models M, one need not compute the denominator, $P(D)$, since it is the same for all candidate models:

$$\hat{M} = \underset{M}{\text{argmax}} \; P(M|D) \tag{5.3.3}$$

$$= \underset{M}{\text{argmax}} \; \frac{P(M) \cdot P(D|M)}{P(D)} \tag{5.3.4}$$

$$= \underset{M}{\text{argmax}} \; P(M) \cdot P(D|M) \quad , \tag{5.3.5}$$

where the $\underset{M}{\text{argmax}}$ operator selects the model M that maximizes the term that follows.

This *maximum a posteriori* (or *MAP*) estimate has found wide applicability in all domains of probabilistic reasoning; the model space M might range over a class of mathematical functions, a set of variables describing a world state, the hidden cause of a medical condition — indeed, any set of competing accounts of some observed data. The current discussion is motivated by the search for linguistic knowledge, where the model space is the space of grammars that can explain a body of linguistic experience.

5.3.2 Bayesian model merging

The closest antecedent to the current enterprise is Bayesian model merging. The basic idea behind model merging (Omohundro 1992) is to treat every encountered instance as an exemplar to be incorporated in its entirety into an (initially) unstructured overall model (in the general sense of *model* noted above). These initial submodels are thus maximally specific, so the model will perform very well on data previously observed (or similar to that previously observed), but poorly on data ranging further afield from its experience. But as more instances are incorporated, the algorithm incrementally modifies the model — in particular, by *merging* submodels — to reflect regularities in the data, as captured by similarities in the submodels. Every merge operation yields a more general model that accounts for a wider range of data.

As with other specific-to-general learning algorithms, the challenge is knowing when to stop:

stopping too early runs the risk of essentially memorizing the data, with the submodels still hewing to specific observed instances; stopping too late may lead to a vacuous model whose powers of discrimination have been compromised. Moreover, at any point, many candidate merge operations may be possible; some of these may strand the model in an inhospitable region of the search space, perhaps even a local minimum.

The Bayesian variant of model merging (Stolcke 1994; Stolcke & Omohundro 1994) addresses these problems by applying the Bayesian MAP criterion to control the merging process, that is, by selecting at every step the merge that leads to the largest increase in the overall probability of the model given the data encountered so far. The MAP estimate provides the means of guiding (and stopping) the search over possible merges. Each merge results in a model that is simpler (*i.e.*, has fewer models, and thus has a higher prior probability) but less specific to the data (*i.e.*, has lower likelihood). The algorithm chooses merges that increase the model's posterior probability, and stops when such merges are no longer available. Stated in its most general form:

Bayesian model merging algorithm (Stolcke 1994; Stolcke & Omohundro 1994)

1. **Data incorporation.** Given a set of examples D, build an initial model M_0 that explicitly includes each example.
2. **Structure merging.** Until the posterior probability decreases:
   (a) Find the candidate merge of substructures that results in the greatest increase in posterior probability.
   (b) Perform the candidate merge and remove the original structures.

The algorithm has a number of properties that make it a particularly appealing and cognitively plausible candidate for the current task. Models at all levels of specificity can happily coexist; aspects of both imitative learning (especially early in learning) and generalization (as more data is encountered) are reflected in model performance, just as in patterns of child language acquisition; and posterior probability provides a principled evaluation metric for guiding the search toward incrementally better models. The algorithm also lends itself to an online version, in which the two steps can be repeated for batches of data (or, in the limit, single examples).

Examples

The following two applications of model merging are especially relevant to the current problem:

Probabilistic attribute grammars. Stolcke (1994) applies model merging to learn several classes of probabilistic language models, including hidden Markov models, probabilistic context-free grammars and probabilistic attribute grammars. The last of these is the most relevant here: a *probabilistic attribute grammar* (PAG) extends a stochastic context-free grammar backbone with probabilistic feature constraints, which can express simple semantic relations. Input is drawn from the L_0 task mentioned in Section 1.2.2 (Feldman *et al.* 1996), based on simple scenes of shapes in a trajector-landmark configuration. Thus sentences like "a circle is above a square" are paired with attribute-value pairs describing the scene (*e.g.*, "tr=circle lm=square rel=above"). The model successfully learns PAGs that parse the input sentence and produce accurate corresponding scene descriptions (analogous to the task of comprehension).

Lexical semantic representations. Bailey (1997) applies the model merging algorithm to a semantically richer though syntactically simpler domain than Stolcke's PAG domain to model the crosslinguistic acquisition of hand action verbs. The semantic domain draws on an active motor representation based on the *x-schema* formalism described in Section 2.3.3. The input consists of single words (*e.g.*, *push* and *shove*) paired with features structures parameterizing x-schemas of the associated actions (*e.g.*, "schema=push force=high posture=palm direction=away ..."). Merging these feature structures yields fewer structures with broader multinomial probability distributions over their features, resulting in a final lexicon containing one or more submodels for each verb. It thus exhibits both polysemy (*e.g.*, with different models for *push*ing a block and *pushing* a button) and near-synonymy (*e.g.*, similar models for *push* and *shove*, where the latter has a higher force component). The model also demonstrates how the same underlying action description, motivated by presumed universals of motor control, can give rise to crosslinguistic diversity in systems for naming actions, as shown for languages including Tamil, Farsi, Spanish and Russian. Learned lexicons allow the model to perform successfully on single-word versions of comprehension" (generation of a feature structure based on a verb) and production (selection of a verb given a feature structure).

These examples demonstrate how the general model merging algorithm can be adapted for a given domain, as summarized below in terms of four main components:

Data incorporation strategy: how to incorporate new data, that is, how to construct the initial model. Typically, a (sub)model is created for each input example, effectively memorizing the data encountered. The PAG case, for example, simply adds a new grammar rule expanding the start symbol into the input sentence and features; while the verb learning model adds a specific verb submodel matching the input verb and feature structure.

Structure merging operations: how to merge substructures. The merging operations typically perform generalization over submodels appropriate to each domain. The verb sense model combines probability distributions over each feature of two merged submodels for a given verb. The PAG model offers more structure (both CFG rules and feature equations) and correspondingly more merging operations. In addition to a generalization operator that merges rules with similar expansions to create more general rules, the model allows *chunking* (or composition) of two nonterminals into a separate new nonterminal, as well as additional operations on the feature equations.

Search strategy: how to search the model space for candidate merges. An exhaustive search through all possible pairs of structures may be possible for small domains, but in practice, heuristics for guiding the search are useful and necessary to reduce computation time. The PAG model and others described in Stolcke (1994) apply best-first search with some lookahead, while the verb sense model uses a *similarity* metric to keep track of the most similar feature structures as the best candidates for merging.

Evaluation criteria: how to compare competing candidates for merging, that is, how to measure the prior and likelihood used to compute posterior probability. In both cases above, the priors bias models toward simplicity (fewer and shorter rules; fewer total verb senses), while the likelihoods are calculated according to the relevant probabilistic model. The PAG model also keeps counts of rule use to avoid costly reanalysis of the entire input data.

Both of these examples are direct predecessors of the current model. The PAG formalism exhibits structural complexity: rewrite rules expand nonterminals to sequences that may include nonterminals that are themselves expanded. This complexity corresponds directly to ECG constituent structure. The verb learning model shows how semantically richer representations can be learned, subject to many of the same cognitive and developmental constraints as the present model.

5.3.3 Model merging for embodied constructions

How can the model merging paradigm be applied to the construction learning scenario at hand? As noted earlier, the domain of embodied constructions differs in important respects from previous language formalisms. A naive translation of the components above will therefore not suffice. Extensions to the model are prompted by three main considerations: the nature of the ECG representation itself; the tight relationship between language learning and understanding; and cognitive constraints on learning.

Representational expressivity. The ECG formalism has explicit notions of category structure, constituency and constraints. Additional structure comes from its basic function of linking the two domains of form and meaning. The model merging algorithm must be modified to handle these sources of structure, both within and across constructions. Operations for modifying the grammar must be extended to accommodate the internal structure of relational constructions, in particular to refine the notion of similarity to recognize the potential for shared, unshared and overlapping structure across constructions. In addition, appropriate evaluation criteria must be defined for scoring ECG grammars, corresponding to the Bayesian prior probability but suitable for the discrete nature of (this version of) the ECG formalism.

Usage-based learning. Processing considerations necessitate fundamental modifications to the model. Unlike previous model merging applications, in which training examples are initially incorporated directly (*i.e.*, memorized) as exemplars, the path from input token to construction is mediated by language comprehension. In fact, the goal of learning is to discover how existing structures can be exploited and combined into larger relational constructions. Thus, it is most economical to take advantage of the partial analyses provided by language comprehension, adding new constructions only as needed to supplement and extend existing constructions. Usage processes should also directly guide the search for grammar modifications: the choice of best operation may be triggered by specific instances of usage as they are analyzed and resolved. Finally, appropriate evaluation criteria must be defined for evaluating how well a grammar performs on a corpus, corresponding to the likelihood. The constructional score defined in Section 4.3 partially fulfills this function, since it measures the constructional probability $P(d|G)$ of the input token d given the grammar G. But the other factors included in Section 4.3 for measuring how well the grammar facilitates language comprehension should also be incorporated, as in Bryant (2008).

Cognitive constraints. Some additional changes to standard model merging are motivated by cognitive considerations, in particular the constraint that the learner may have limited storage memory capacity for both input tokens and constructions, as well as limited working memory to support the learning process. Incorporating each new example as a separate model, for example, may be reasonable for (some) lexical items and short multi-word sequences, but automatically codifying every new input token in its full glory as a novel construction seems less cognitively plausible. It may also be infeasible to search through the entire space of possible structure merging operations at every step, or to assess how a candidate construction affects a grammar's posterior probability by reevaluating the entire corpus of previously seen data. Appropriate heuristics for limiting the search space or reducing computation are needed; these will be discussed in Chapter 7..

5.4 Usage-based construction learning: overview

This section synthesizes the considerations above to present a usage-based model of construction learning that addresses the general class of language learning problems defined in Section 5.2. Figure 5.2 revisits the integrated understanding-learning cycle introduced in Chapter 1 (Figure 1.1). The structures and processes involved in the language understanding pathway — including embodied schemas and constructions, the input token of an utterance in its situational context, the analysis and resolution processes and the resulting semspec — have now been explicitly defined.

The language understanding process provides a direct usage-based impetus for the language learning part of the cycle, depicted here as two processes: the *hypothesis* of new constructions and the *reorganization* of existing ones. These construction learning operations, corresponding to the structure merging operations in model merging, are motivated and constrained by the nature of the target space, the processes of language comprehension and cognitive considerations. They are also are mediated by quantitative evaluation metrics (not shown in the figure), approximating Bayesian scoring criteria. As the learner encounters more data in the usage-learning loop, the constructional mappings learned and refined should facilitate increasingly more complete and accurate comprehension of new data.

Figure 5.3 gives a high-level incremental learning algorithm corresponding to Figure 5.2, along with a version taking a corpus of input tokens that simply loops over the above procedure. The algorithm provides a class of solutions to the language learning problems defined in Section 5.2, adapting Bayesian model merging for the construction-learning domain to address the concerns outlined in Section 5.3.3 as follows:

Data incorporation: Input tokens are not incorporated directly as new constructions, but are instead first analyzed and resolved using the current grammar. They are indirectly incorporated via learning operations that propose new constructions based on the results of analysis.

Grammar update operations: Structure merging operations are extended to allow both the hypothesis of new constructions and the reorganization of existing ones. These must satisfy both the structural demands of the ECG construction formalism (to exploit the presence of shared internal structure) and the process-based demands of language analysis (to improve comprehension of new input).

Search strategy: The search for candidate constructions is guided by incoming data. In some cases, the results of analyzing the current input token directly triggers specific learning operations.

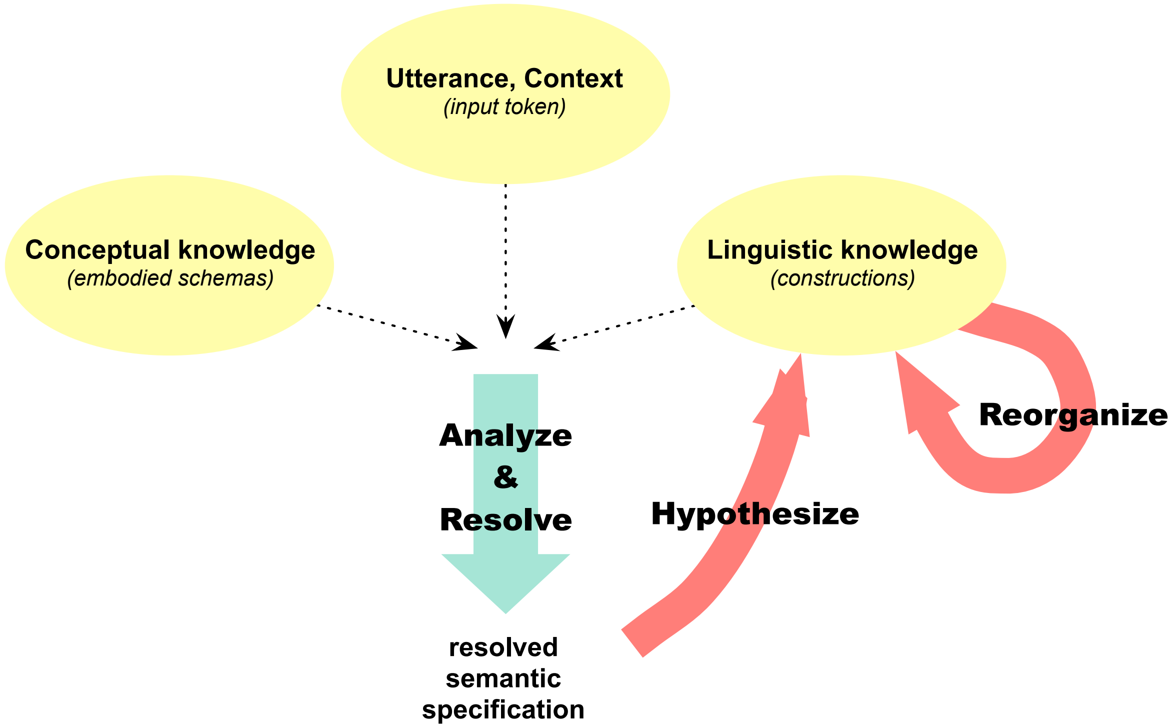

Figure 5.2. Usage-based construction learning: Input tokens (utterance-context pairs) are analyzed and resolved, producing a resolved semantic specification and prompting the hypothesis of new constructions and the reorganization of existing ones.

Learn from input token: Given input token d and grammar G, return grammar G'.

1. Analyze d using G, producing analysis a_d and resolved semspec rss_a.
2. Find the set L of candidate learning operations based on a_d, rss_a and G, by hypothesizing new constructions and reorganizing existing constructions.
3. For each learning operation $l \in L$:
   (a) Create new grammar G_l by performing operation l on G.
   (b) Calculate the grammar improvement $\Delta(G_l, G)$.
4. Set G' to the G_l that maximizes $\Delta(G_l, G)$.
5. If $\Delta(G', G)$ exceeds threshold of improvement, return G'; else return G.

Learn from corpus: Given a corpus D and a grammar G_0, return grammar G_1.

1. Initialize current grammar G to G_0.
2. For each input token $d \in D$, learn from d and current grammar G, producing G'.
3. Return current grammar as G_1.

Figure 5.3. Usage-based construction learning algorithm: algorithms for learning new grammars from a single token and from a corpus.

In others, the search for new constructions may be restricted to a subset of the entire repository (*e.g.*, recently used constructions and their nearest neighbors in the construction set).

Evaluation criteria: The grammar improvement $\Delta(G', G)$ corresponds to the posterior probability used in model merging, but it is adapted for the discrete structures of the ECG domain to use a simplicity-based criterion that balances a prior favoring simpler grammars against a likelihood favoring simpler analyses of the data using the grammar (*i.e.*, encoding a bias toward better prediction of the data). The calculation of this score may be restricted to a subset of the corpus (*e.g.*, recently encountered tokens, or especially relevant or problematic tokens).

Besides the representational modifications needed for the ECG domain, the most significant adaptations to each component above are motivated by aspects of language comprehension. In particular, the shift to a data-driven, analyzer-mediated basis for processing input removes the strict boundary between data incorporation and the search for new constructions: both depend on the ongoing flow of data as it is processed by the current grammar. The model is thus usage-based in that its path through the space of grammars depends directly on specific instances of language use. It is also usage-based in the related sense of exploiting the statistical characteristics of usage over time, as reflected in the model's scoring criterion.

Like the general class of problems it addresses, this class of solutions admits many instantiations; each of the algorithmic components above leaves significant room for interpretation. The next two chapters together instantiate a solution that satisfies the constraints of the construction learning problem defined in this chapter, where the ECG-based structures described in Chapter 3 and, especially, the language analysis processes described in Chapter 4 bear directly on every aspect of the model.

- Chapter 6 addresses issues related to the search for candidate constructions, motivated by both the nature of the search space and domain-specific strategies suggested by developmental evidence. I define several operations for the hypothesis and reorganization of constructions and describe the conditions under which they apply (corresponding to step 2 of the high-level algorithm).

- Chapter 7 focuses on the quantitative evaluation of candidate grammars, defining a heuristic for calculating the grammar improvement $\Delta(G', G)$ based on minimum description length. This heuristic is guided by an information-theoretic bias toward a minimal encoding of the grammar together with the data.

In fleshing out the requisite details, I will make a number of simplifying assumptions to ease both exposition and implementation. It is essential to bear in mind that the intent is not to define an exhaustive set of learning operations or the most statistically sensitive evaluation criteria possible. The nature of the problem space and relative paucity of data available for experimentation make it unlikely to reward much ingenuity in model design at this initial stage. Fortunately, the learning framework guarantees that with enough data, any reasonable optimization criteria should lead to improvement over time. The goal here is thus to define a basic toolkit of usage-based operations and criteria, sufficient to demonstrate how the underlying dynamics of the model push it toward increasingly better grammars over the course of experience.

Chapter 6

Usage-based learning operations

Experience is not what happens to you;
it is what you do with what happens to you.

— *Aldous Huxley*

How might a language learner channel experience into knowledge? The model laid out in the last chapter requires both a search through the space of possible grammars and a evaluation of candidates. The latter of these, the quantitative evaluation, is specified by the general learning framework to depend on maximizing posterior probability, or some approximation thereof (discussed in Chapter 7). This framework is mute, however, on the question of how to explore the hypothesis space. For this the model relies on domain-specific heuristics, along with the overarching constraint that the operations be usage-based: the learner is assumed to be not just passively receiving input tokens but actively trying to do something with them, *i.e.*, make sense of them using the language understanding processes described in Chapter 4.

Section 6.1 motivates and summarizes several kinds of learning operations available to the model; these are elaborated in Section 6.2.2 (mapping operations) and Section 6.3 (reorganization operations). Section 6.4 relates these to other learning techniques proposed in the computational and psychological literature.

6.1 Overview of operations

The current model is concerned primarily with the learning of relational constructions, *i.e.*, structured mappings across the domains of form and meaning. Such mappings are motivated, initially, by the latent structure of experience: both utterances and their surrounding contexts are rife with entities (forms and meanings, respectively) and relations among them. The learner's task is to extract useful patterns (*i.e.*, new constructions) from this raw stream of complex relationships. The learner may also consolidate or otherwise reorganize existing constructions. These two sources of new constructions — the input data, and the current grammar — correspond to two broad classes of learning operations:

map forms and meanings co-occurring in an utterance-context pair

- *Simple map.* Hypothesize a simple map between a set of forms and a set of meanings.
- *Relational map.* Hypothesize a structured map over the forms and meanings of known constructions.

reorganize existing constructions to exploit shared structure

- *Merge.* Create a (generalized) construction from structure shared by two constructions.
- *Join.* Create a new construction combining structure from two constructions.
- *Split.* Create a new construction by extracting structure present in one construction but not another.

This division of operations highlights the parallels between this task and the previous applications of model merging described in Section 5.3.2. Context-driven mapping operations serve as the primary route from input data to new structural relations, and are thus analogous to the *data incorporation* phase. These vary in the the extent to which they exploit existing constructions found in a partial analysis of an input token, but in all cases they encode aspects of a co-occurring utterance and context (Figure 6.1).

Abstracting from the particulars of form and meaning, we can also view constructions as discrete chunks of information that may overlap with each other. The reorganization operations exploit this shared structure to reshuffle these informational chunks into new combinations, using constructional analogues of standard set-based intersection, union and difference (corresponding to the Venn diagrams in Figure 6.2). These operations correspond to the *structure merging* phase of model merging, though they expand the set of possibilities beyond those used in the simpler domains previously described.

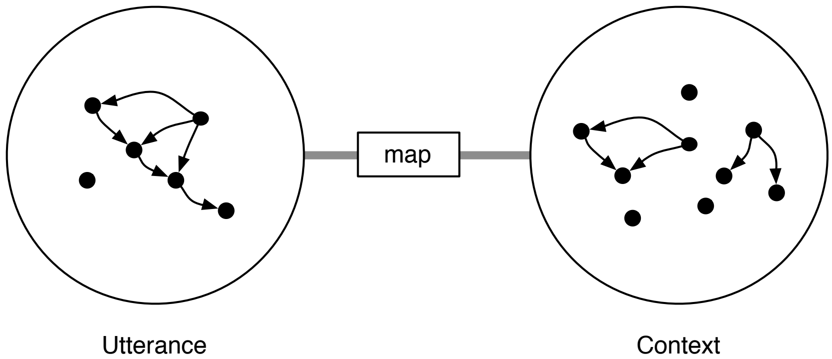

Figure 6.1. Mapping an utterance to context

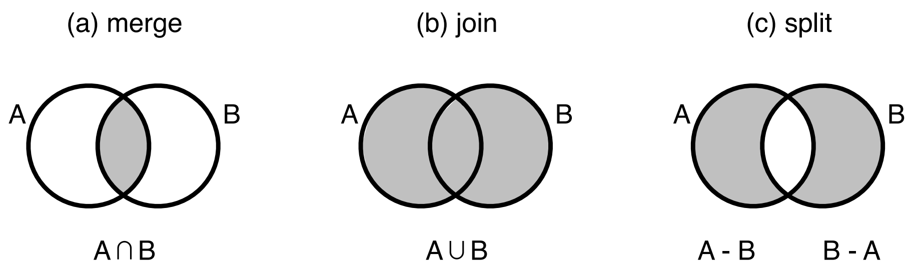

Figure 6.2. Exploiting shared structure: (a) merge based on shared content; (b) join based on shared and unshared content; and (c) split based on unshared content.

Together, these operations provide well-defined mechanisms by which the learner can traverse the candidate grammars reachable from the current grammar, based on the current input. This set, though finite, remains dauntingly large. Thus, while the quantitative criteria described in Chapter 7 remain the ultimate means for evaluating potential grammars, the operations themselves can also be constrained to limit the search space in various ways — for example, to prefer constructions that hew more closely to the observed input, or that draw on cognitively motivated, domain-general learning strategies. The sections to follow include both general descriptions of each operation as well as heuristics for restricting them by exploiting such usage-based considerations.

6.2 Context-driven mapping operations

On the constructional view, linguistic knowledge reflects entrenched patterns of experience: spatio-temporal co-occurrence of specific forms and meanings is taken to be coincidental only in a literal sense, and repeated such encounters reinforce these associations. This kind of learning naturally extends the Hebbian principle: forms and meanings that fire together, wire together. The resulting associations provide the basis for the emergence of communicative functions (*e.g.*, naming, referring, predicating).

One approach to capturing such correlations is to directly memorize input tokens as they are observed, as suggested by the utterance-context mapping shown in Figure 6.1. That is, the specific form and meaning information observed as part of an input token can be effectively incorporated as a maximally specific construction. Such a procedure may be at work for many varieties of word learning, based on the pairing of an observed (single) form with some referent object or scene.[1] Fixed multiword expressions may be similarly straightforward to model as rote learning. More generally, however, the input token may pair a multiword utterance with a rich contextual environment, making it infeasible to retain the entirety of the input token's information. Form-meaning mappings can still, in this more complex scenario, be lifted from experience, but the learner may seek a subset of the utterance to associate with a subset of the context.

The context-driven mapping operations below all exploit the language comprehension processes described in Chapter 4 to choose a reasonable subset of the utterance and context to map. These processes effectively partition the input forms and meanings according to whether they are accounted for by the current grammar; this information directly prompts the formation of new form-meaning mappings. Strategies for hypothesizing both simple and structured constructions are considered below.

6.2.1 Simple mapping

The catalogue of constructional mapping types given in Figure 4.2 includes several simple maps between the form and meaning domains (a-c), all characterized by the lack of intervening structure in the constructional domain. Such maps are canonically most closely associated with lexical items and holophrases (*i.e.*, constructions whose forms are typically treated as unanalyzed). Though lexical learning is not the primary focus of the current work, some of the same strategies useful for restricting the set of potential mappings also apply to relational learning strategies.

In both cases, an incomplete (or incorrect) grammar may yield a partial analysis whose associated semspec is imperfectly resolved against the current context. The resulting discrepancies — *i.e.*, the unmatched forms and meanings that are presumably not yet explained by known construc-

[1] As discussed in Section 2.1.3, child language learners may draw upon sophisticated pragmatic inference skills to infer the appropriate referent for a word. Once inferred, the association does not require prolonged exposure to stabilize, as suggested by evidence of fast mapping (Carey 1978; P. Bloom 2000).

tions — provide obvious candidates for learning. The simple mapping algorithm below includes several heuristics for selecting appropriate subsets of the analysis A:

Simple map: Given analysis A of input token $d = (u, z)$, return a simple construction $\mathbf{c}$.

1. Find the sets $\bar{A}_f$ and $\bar{A}_m$ of unmatched forms and meanings.
2. Create a new construction $\mathbf{c}$, with $\mathbf{c}_f \subseteq u$ and $\mathbf{c}_m \subseteq z$. Mapping strategies include:
   - Map all: Set $\mathbf{c}_f$ = u and $\mathbf{c}_m$ = z.
   - Map unmatched: Set $\mathbf{c}_f = \bar{A}_f$ and $\mathbf{c}_m = \bar{A}_m$.
   - Map unmatched subset: Set $\mathbf{c}_f = f \subset \bar{A}_f$ and $\mathbf{c}_m = m \subset \bar{A}_m$.

The heuristics above could be refined to incorporate developmentally motivated preferences, such as a preference for forms and meanings that are *salient* (*e.g.*, stressed words, or contextual items that are in attentional focus), or any of the various other principles of word learning mentioned in Section 2.1.3. Other heuristics proposed in the literature are noted in Section 6.4.[2] All of these are compatible with the current framework and could be used in conjunction with the structure-building operations that are its main focus.

In the experiments to be described, the model reserves simple mapping for unambiguous situations in which a single unfamiliar word in the utterance appears with a single salient unmapped context item. (This operation is equivalent to one kind of splitting operation, as discussed in Section 6.3.4.) More concretely, given an unknown word $f \in \bar{A}_f$ and an unmatched context item $m \in \bar{A}_m$, the model creates a simple lexical construction $\mathbf{c}_f$ whose form pole is a Word with its orthography role set to the observed orthography of f and whose meaning pole is typed as the observed type of m. (The meanings are thus lifted from a particular concrete context item into the corresponding meaning schema.) The resulting constructions resemble the lexical items assumed in the initial lexicon and exemplified in Figure 4.3.

6.2.2 Relational mapping

Relational mapping is the most representationally potent of the proposed model's learning operations: the hypothesis of structured mappings of the kind depicted in Figure 4.2(d) is what gets cross-domain constituent structure off the ground. All relational constructions are, in the current model, either the direct result of relational mapping or derived through reorganization operations applied to the results of relational mapping.

[2]Many computational models of lexical acquisition (including some of those mentioned in Section 2.3.3, and the Bailey model for verb learning discussed in Section 5.3.2), can also be cast as building initial models by applying the "map all" strategy, where the input utterance consists of a single word.

The operation exploits the partial comprehension as described above for simple mapping, with an additional heuristic that favors relational form-meaning correspondences. The motivating intuition is that structure in one domain may be more easily learned if it is coordinated with structure in another domain — whether because it is more salient to identify that structure, or because encoding both structures together is more concise than encoding them in isolation. Conversely, complex relations that have no shared structure are more difficult to discern and maintain.

Mapping is best illustrated by a simple example, based on analyzing the input token in Figure 5.1. Assume a grammar based on G_1 (defined in Chapter 4) but containing only lexical constructions (*i.e.*, without the THROW-IMP construction), as in the lexical-only analysis discussed in Section 4.3.3.[3] The semspec produced by analysis of the utterance "throw the ball" includes the Ball and Throw schemas, without indicating any relations between them. But both are matched during resolution to the contextually available items; the resulting resolved semspec is graphically depicted in Figure 6.3, where the two lexical constructions matched by the analyzer are shown in the center. Items drawn with with heavy lines correspond to the result of constructional analysis; the rest are present in the input but unmatched by analysis, *i.e.*, they are part of $\bar{A}$, and not predicted by linguistic knowledge. (The unmatched form *the* is not shown.)

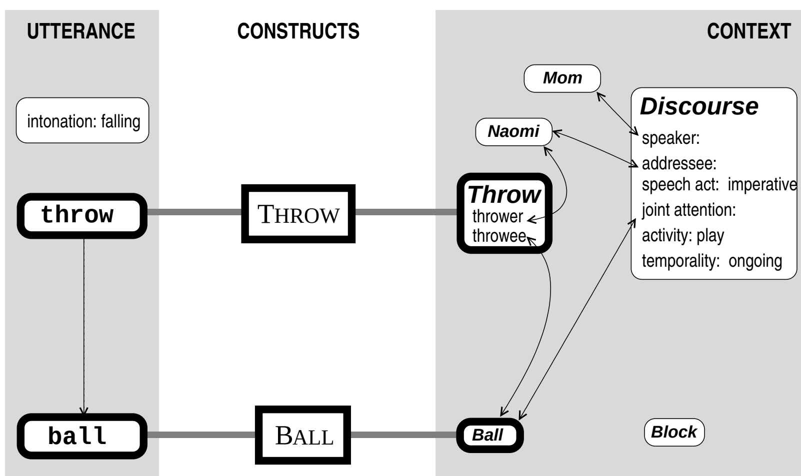

Figure 6.3. Potential mappings in a resolved analysis based on input token *throw the ball*. Heavy solid lines indicate structures matched during analysis and resolved to items shown in context. Relations connecting these (*i.e.*, one that are available in the input but unmatched by analysis) are ideal candidates for new constructional mappings.

As in simple mapping, a new construction might be hypothesized to account for the input, perhaps focusing on the unmatched subset of forms and meanings. In particular, the subset can be

[3]Recall that the token analyzed in Section 4.3.3 differs from that in Figure 5.1 only in that its context includes a Throw schema. As noted in Section 4.1.2, this reflects the child's presumed ability to infer her mother's communicative intent.

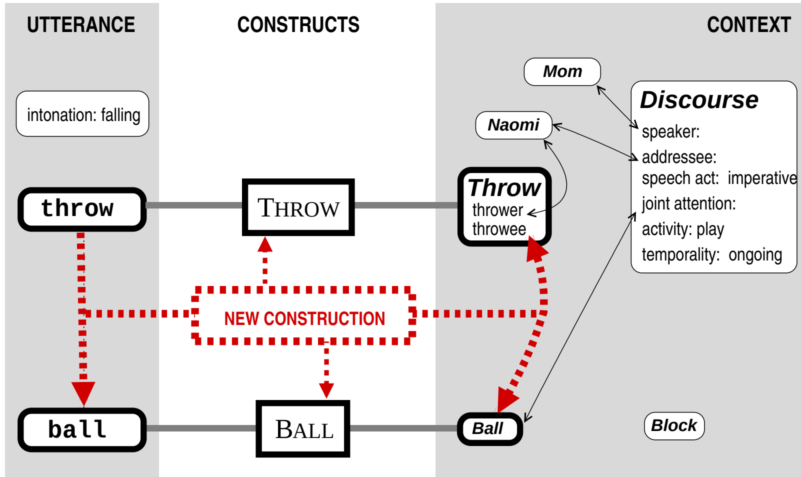

Figure 6.4. Hypothesizing a relational mapping for the utterance *throw the ball*. Heavy solid lines indicate structures matched during analysis; heavy dotted lines indicate the newly hypothesized mapping, between an unmatched ordering relation and role-filler binding.

chosen to exploit the presence of the two matched lexical constructs — that is, one notion of *salience* might be defined to favor schemas and relations that are closely connected to those matched by the analysis. In the example, both domains include unmatched relations that link the domain poles of the respective constructs: the form domain has a word order edge between *throw* and *ball*, while the meaning domain includes a role-filler binding between the Throw.throwee role and the specific Ball in context. A few other candidate meaning relations, such as those linking discourse information with other contextually available items, are depicted in the figure.

In this case, the relations linking constructionally mapped items in each domain might be considered the most salient; these relations are drawn in Figure 6.4 with heavy dashed lines. The key observation is that the two mapped relations have *isomorphic* structure, serving as the connection between the forms and meanings of the already known lexical items. A hypothesized (lexically specific) construction motivated by this structural correspondence is shown in Figure 6.5. The construction has two constituents, named t1 and t2 and typed THROW and BALL, respectively.[4] The two relations in the figure are captured as the form constraint $t1_f$ before $t2_f$ and the meaning constraint $t1_m$.throwee $\longleftrightarrow$ $t2_m$, where each constraint element is written in terms of the shortest path (or slot chain) from the relevant constituent.[5]

The structural correspondence above suggests a *relational mapping* heuristic for hypothesizing a

[4]These are the same as the THROW-CN and BALL-CN constructions defined in Chapter 4; the simplified names are used in this chapter to ease exposition.

[5]As this example suggests, two relations may have isomorphic structure despite differing formal notations. That is, the ECG formalism notates word order relations and schema role-filler bindings differently, but their graph-theoretic interpretation is structurally equivalent.

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

Figure 6.5. Construction resulting from relational mapping

new construction, where candidate constructions must satisfy either **strict isomorphism** or a more permissive structural correspondence of **pseudo-isomorphism**, defined below:

Given domain graphs F and M, constructs a and b, a path $r_f(a_f, b_f)$ between nodes $a_f, b_f \in F$ and a path $r_m(a_m, b_m)$ between nodes $a_m, b_m \in M$:

- r_f and r_m are **strictly isomorphic** if $|r_f| = |r_m| = 1$, *i.e.*, r_f and r_m each consist of a single edge linking the respective poles of a and b.
- r_f and r_m are **pseudo-isomorphic** up to distance_p if $|r_f| + |r_m| \leq \text{distance}_p$.

This condition enforces structural correspondences in the form and meaning domains while recognizing that constructions may involve relations that are not strictly isomorphic. The limit distance_p puts an upper bound on the total path length; a setting of $\text{distance}_p = 2$ corresponds to strictly isomorphic relations, while $\text{distance}_p = 3$ corresponds to the most restrictive pseudo-isomorphism, in which the path in one domain is a single edge and in the other is a path of length 2, *i.e.*, one connecting two nodes with a single intervening node. More permissive settings of distance_p allow more intervening nodes between mapped elements, thus increasing the set of candidate mappings.[6] The current model sets the pseudo-isomorphic bound to $\text{distance}_p = 3$ and further requires that $|r_f| = 1$ (*i.e.*, two forms must be directed related[7]). This asymmetry reflects the relatively richer structure available in the meaning domain for the learning problem faced by the child, though in principle the extra structural flexibility could reside in either domain.

Figure 6.6 depicts three common patterns (along with examples) of relational form-meaning mappings, *i.e.*, ways in which a meaning relation over a_m and b_m can be correlated with a form relation over a_f and b_f:

(a) *strictly isomorphic:* b_m is a role-filler of a_m, or vice versa (a_m.r $\longleftrightarrow$ 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)

The relational mapping algorithm suggested by the foregoing discussion is defined below:

[6]Relaxing the isomorphic condition may be desirable if the current input token offers no mappings that satisfy the default condition. Since these more indirect relations are expressed by longer ECG constraints, however, they are penalized by the evaluation criteria described in Chapter 7.

[7]Recall, however, that the current model includes both the contiguous `meets` and non-contiguous `before` as possible form relations; thus two forms may be non-contiguous yet directly (graph-)related in the sense relevant here.

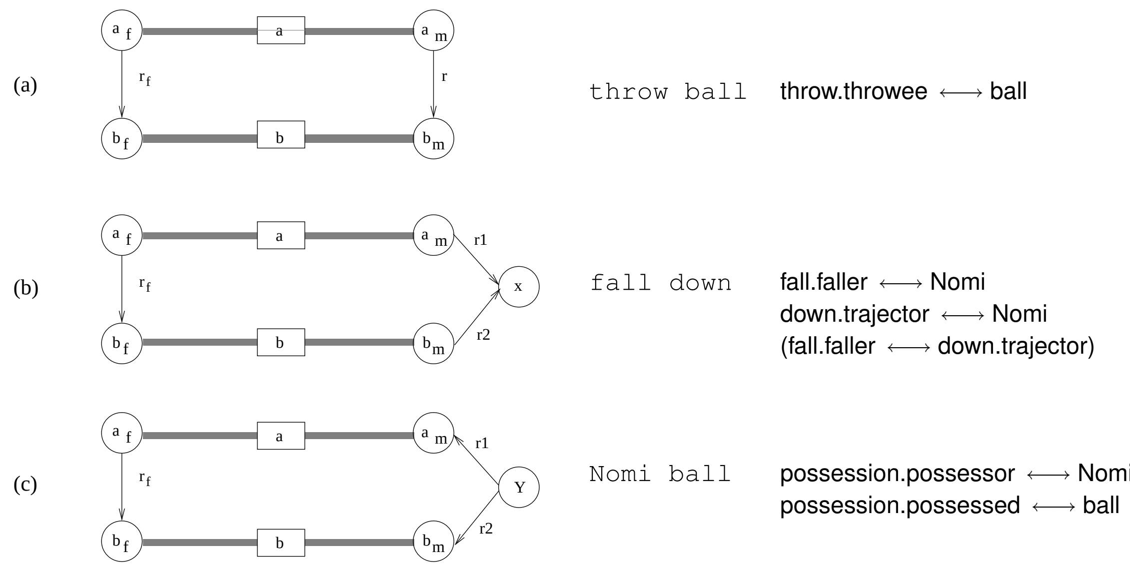

Figure 6.6. Pseudo-isomorphic relational mappings over constructs a and b: (a) strictly isomorphic; (b) shared role-filler; and (c) sibling role-fillers.

Relational map: Given analysis A of input token $d = (u, z)$, return a relational construction $\mathbf{c}$.

1. Find the sets $\bar{A}_f$ and $\bar{A}_m$ of unmatched forms and meanings.
2. Choose constructs a and b in $\text{constructs}(A)$ that are structurally isomorphic:
   - a_f and b_f are connected by a path $r_f(a_f, b_f) \subset \bar{A}_f$;
   - a_m and b_m are connected by a path $r_m(a_m, b_m) \subset \bar{A}_m$; and
   - $r_f(a_f, b_f)$ and $r_m(a_m, b_m)$ are strictly isomorphic or pseudo-isomorphic.
3. Create a new construction $\mathbf{c}$, with:
   - $\text{constituents}(\mathbf{c}) = \{n1, n2\}$;
   - $\text{type}(n1) = \text{type}(a)$ and $\text{type}(n2) = \text{type}(b)$; and
   - $r_f(a_f, b_f) \in \text{constraints}(\mathbf{c}_f)$ and $r_m(a_m, b_m) \in \text{constraints}(\mathbf{c}_m)$,

   where r_f and r_m are rewritten in terms of the poles of the constituents $n1$ and $n2$.

Like simple mapping, the algorithm first limits the search for candidate relations to those unmatched by the current analysis but present in the situation. It then seeks relations that capture (pseudo-)isomorphic structure across the form and meaning domains. In effect, the algorithm searches for heretofore unpredicted relations that may be considered the most *relevant* and *salient* in the current situation, by virtue of being anchored by known constructions instantiated in the analysis.[8] Suitable relations are then encoded as the form and meaning constraints, respectively, of a new construction whose constituents correspond to (and have the same type as) the constructs anchoring the new relations; constraints are rewritten (as above) in terms of the newly created con-

[8]The algorithm as shown allows only two such constructions at a time, though in principle a larger number of constructions is possible.

stituents. This hypothesized construction is consistent with the current input token; a new grammar including it explains more of the form and meaning constraints than the current grammar.

More complex examples offer many possible mappings satisfying the relational mapping conditions. The current model encodes a preference for shorter relational paths in two ways: (1) by requiring strictly isomorphic ($\mathrm{distance}_p = 2$) or minimally pseudo-isomorphic relations ($\mathrm{distance}_p = 3$); and (2) by allowing relations only on the root constructs of the analysis.

As with simple mapping, alternative heuristics could allow the mapping of all matched or unmatched forms and meanings, or choose a subset based on some other criteria, *e.g.*, a preference for some privileged subset of salient entities or relations. The current model includes, in addition to the structural isomorphic preference, a preference (in the meaning domain) for *discourse* mappings that involve roles of the current discourse space. These heuristics are illustrated in action in the experimental results of Chapter 8.

6.3 Reorganization operations

This section considers a second class of operations, those that infer new constructions by reorganizing the existing set of constructions. In particular, shared structure across multiple constructions may suggest potential areas of consolidation, which in turn may reveal useful general mapping patterns. They are thus analogous to the merging operations used in previous applications of model merging discussed in Section 5.3.2, but they are extended to accommodate the complex, relational structures needed for multiword constructions. In addition to a merge (or generalization) operation based on constructional intersection (merge), operations corresponding to constructional union (join) and difference (split) are introduced.

Unlike the mapping operations just discussed, reorganization operations do not require explicit reference to the current context; rather, they act only on the set of known constructions, and could thus be conceptualized as proceeding in parallel or interleaved with the mapping operations. But as noted in Section 5.4, the two kinds of operations need not be strictly separated. In fact, there are both computational and cognitive reasons to apply heuristics to focus the search for candidate reorganization operations around the incoming stream of input tokens, thereby reducing the space of options to consider at any step. Some suitable heuristics for doing so are noted below.

6.3.1 Structural alignment

All of the reorganization operations described in this section rely on the identification of shared structure across constructions. Two constructions may, for example, include constituents of the same (or related) constructional types, and form and meaning poles of the same (or related) schema types; they may also assert similar form and meaning constraints. To identify how and to what extent two constructions overlap in this manner, the model employs a **constructional alignment** procedure that finds the best **constituent map** between two constructions (a set of mappings between compatible constituents) and its best associated **constraint map** (a set of mappings between compatible constraints).

The goal of constructional alignment is to determine whether two constructions that have no *explicitly* defined type relation in the current grammar might be **implicitly** type-compatible — *i.e.*, whether one construction can be deemed implicitly equivalent to ($\approx$), more specific than ($\prec$), more general than ($\succ$), indirectly related to ($\sim$) or incomparable with ($\perp$) another. The determination of this implicit type relation depends in part on the explicit type relations previously defined in Section 4.1.1 for the construction and schema sets; these are extended in Figure 6.7 to allow types (for both constructions and schemas) to be either *directly related* (one is a subcase of the other) or *indirectly related* (through a shared ancestor).

The inherently bipolar nature of constructions creates potential for type comparisons across multiple domains: implicit type relations between two constructions depend in part on the (explicit) type relations between their form and meaning poles, captured by the $\mathrm{compare}_p$ pole-type comparison function defined in Figure 6.7. That is, two constructions that are not explicitly defined as type-related may be implicitly related on the basis of having (consistently) related pole types.

Besides pole-based relations, constructional alignment must also take into account other areas of potential overlap. As defined in Figure 6.8, the overall (implicit) type relation $\mathrm{compare}$ between two constructions depends on specific mappings establishing how (and whether) their constituents and constraints are correlated. Both the constituent map m_n and the constraint map m_r have implicit type comparison values ($\mathrm{compare}_n(m_n)$ and $\mathrm{compare}_r(m_r)$, respectively); these functions collectively determine for a particular alignment of constituents and constraints whether there is a consistent equivalence or subsumption relationship between the two constructions. For constituents, this value is a composite over all its mapped constituents ensuring consistent directionality of subsumption relations ($\prec$ or $\succ$), if applicable. For other constraints (*i.e.*, besides type-specifying dec-

Given elements x, $y \in$ schema set S or construction set C:

- $\mathrm{lca}(x, y)$ is the least common ancestor of x and y.
- Elements x and y are **directly related** if one of them is an ancestor of the other, *i.e.*, if $x \leq y$ or $y \leq x$. In this case, $\mathrm{lca}(x, y) = x$ or $\mathrm{lca}(x, y) = y$.
- Elements x and y are **indirectly related** if they are not directly related but have a common ancestor, *i.e.*, if $\mathrm{lca}(x, y) = z$, $z \neq \top, z \neq x, z \neq y$. This relation is also notated $x \sim y$.
- Elements x and y are **type-compatible** if they are either directly or indirectly related.
- Elements x and y are **incomparable** if they have no common ancestor, *i.e.*, if $\mathrm{lca}(x, y) = \top$. This relation is also notated $x \perp y$.
- $\mathrm{distance}(x, y)$ is the length of the shortest path between x and y.

Given constructions $\mathbf{a}$, $\mathbf{b} \in$ construction set C:

- The domain-specific least common ancestors are defined as $\mathrm{lca}_f(\mathbf{a}, \mathbf{b})$ = $\mathrm{lca}(\mathbf{a}_f, \mathbf{b}_f)$ and $\mathrm{lca}_m(\mathbf{a}, \mathbf{b})$ = $\mathrm{lca}(\mathbf{a}_m, \mathbf{b}_m)$.
- The domain-specific distance functions are $\mathrm{distance}_f(\mathbf{a}, \mathbf{b})$ and $\mathrm{distance}_m(\mathbf{a}, \mathbf{b})$.
- $\mathrm{compare}_p(\mathbf{a}, \mathbf{b})$ is an *implicit* type comparison based on pole types, defined as:

  $\approx$ (equivalent), if $\mathrm{type}(\mathbf{a}_f) = \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) = \mathrm{type}(\mathbf{b}_m)$

  $\prec$ (more specific), if $\mathrm{type}(\mathbf{a}_f) \leq \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) \leq \mathrm{type}(\mathbf{b}_m)$,

  $\succ$ (more general), if $\mathrm{type}(\mathbf{a}_f) \geq \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) \geq \mathrm{type}(\mathbf{b}_m)$,

  $\sim$ (indirectly related), if $\mathrm{type}(\mathbf{a}_f) \not\perp \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) \not\perp \mathrm{type}(\mathbf{b}_m)$, *and*
  (($\mathrm{type}(\mathbf{a}_f) \sim \mathrm{type}(\mathbf{b}_f)$ or $\mathrm{type}(\mathbf{a}_m) \sim \mathrm{type}(\mathbf{b}_m)$) or
  ($\mathrm{type}(\mathbf{a}_f) < \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) > \mathrm{type}(\mathbf{b}_m)$) or
  ($\mathrm{type}(\mathbf{a}_f) > \mathrm{type}(\mathbf{b}_f)$ and $\mathrm{type}(\mathbf{a}_m) < \mathrm{type}(\mathbf{b}_m)$))

  $\perp$ (incomparable) otherwise, *i.e.*, if $\mathrm{type}(\mathbf{a}_f) \perp \mathrm{type}(\mathbf{b}_f)$ or $\mathrm{type}(\mathbf{a}_m) \perp \mathrm{type}(\mathbf{b}_m)$.

Figure 6.7. Extended ECG type relation definitions

larations), the type comparison depends mainly on whether one construction's constraints are a subset of the other's.[9]

All of this machinery is designed to pinpoint areas of structural overlap that may suggest ways in which the grammar can be consolidated. It thus adapts the notion of similarity to accommodate the structural complexities of the constructional domain, specifically as presented by constituency and relational constraints. Additional complications arise from the need to evaluate multiple alignment possibilities. The model employs ranking heuristics that prefer constituent maps that minimize the type distance across mapped constituents (including constructional, form and meaning distances); and constraint maps that maximize the number of shared constraints.

It may be costly in general to perform constructional alignment between all possible pairs of constructions in the grammar, even if the model need do so only periodically. The search can be restricted to permit only alignments within a specified maximal alignment distance, $\mathrm{distance}_a$. Setting $\mathrm{distance}_a$ to 2, for instance, requires mapped constituents to be of the same type, in a direct subcase

[9]The current model has a limited type hierarchy on constraint types, where the word order relation before is subsumed by the more general meets relation. Identity and filler constraints are evaluated only as equivalent (under some rewriting) or not equivalent.

A **constructional alignment** $\mathrm{alignment}(\mathbf{a},\mathbf{b})$ between constructions $\mathbf{a}$ and $\mathbf{b}$ consists of a constituent map $m_n = \mathrm{cmap}(\mathbf{a},\mathbf{b})$ and a constraint map $m_r = \mathrm{ctmap}(\mathbf{a},\mathbf{b},m)$.

- $\mathrm{compare}(\mathbf{a},\mathbf{b})$ is the composite type comparison over m_n, m_r and $c_p = \mathrm{compare}_p(\mathbf{a},\mathbf{b})$, defined as:

  $\approx$, if $\mathrm{compare}_n(m_n)$, $\mathrm{compare}_r(m_r)$ and c_p are all $\approx$

  $\prec$, if $\mathrm{compare}_n(m_n)$, $\mathrm{compare}_r(m_r)$ and c_p are all either $\prec$ or $\approx$

  $\succ$, if $\mathrm{compare}_n(m_n)$, $\mathrm{compare}_r(m_r)$ and c_p are all either $\succ$ or $\approx$

  $\sim$, if none of $\mathrm{compare}_n(m_n)$, $\mathrm{compare}_r(m_r)$ and c_p is $\perp$, *and* at least one is $\sim$, or their values $\in \{ \prec, \succ \}$ conflict

  $\perp$ otherwise, *i.e.*, if any of $\mathrm{compare}_n(m_n)$, $\mathrm{compare}_r(m_r)$ or c_p is $\perp$
- $\mathbf{a}$ and $\mathbf{b}$ have *global shared structure* if there is at least one total constituent map.
- $\mathbf{a}$ and $\mathbf{b}$ are *mergeable* if $\mathrm{compare}(\mathbf{a},\mathbf{b}) \neq \perp$.
- $\mathbf{a}$ and $\mathbf{b}$ have **overlapping structure** if there is at least one constituent map m such that $|m| > 0$.

A **constituent map** $\mathrm{cmap}(\mathbf{a},\mathbf{b})$ between constructions $\mathbf{a}$ and $\mathbf{b}$ is a set of pairs $\langle n_a, n_b\rangle \subset \mathrm{constituents}(\mathbf{a}) \times \mathrm{constituents}(\mathbf{b})$.

- The size of a constituent map is $|\,\mathrm{cmap}(\mathbf{a},\mathbf{b})|$. A constituent map is *total* if $|\,\mathrm{cmap}(\mathbf{a},\mathbf{b})| = |\,\mathrm{constituents}(\mathbf{a})| = |\,\mathrm{constituents}(\mathbf{b})|$, and *partial* otherwise.
- For all $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$, n_a and n_b are directly or indirectly related. That is, $\mathrm{compare}_n(\langle n_a, n_b\rangle) \in \{ <, >, =, \sim \}$.
- $\mathrm{compare}_n(\mathrm{cmap}(\mathbf{a},\mathbf{b}))$ is the composite constituent type comparison, defined for total constituent maps as follows:

  $\approx$, if $n_a = n_b$ for all $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$

  $\prec$, if $n_a \leq n_b$ for all $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$

  $\succ$, if $n_a \geq n_b$ for all $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$

  $\sim$ otherwise, *i.e.*, if $n_a \sim n_b$ for any $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$, or else if there are multiple conflicting (explicit) type relations over all $\langle n_a, n_b\rangle \in \mathrm{cmap}(\mathbf{a},\mathbf{b})$

  $\mathrm{compare}_n(\mathrm{cmap}(\mathbf{a},\mathbf{b}))$ is defined as $\perp$ for partial constituent maps.

A **constraint map** $\mathrm{ctmap}(\mathbf{a},\mathbf{b},m)$ associated with constituent map m_n is a set of pairs $\langle r_a, r_b\rangle \subset \mathrm{constraints}(\mathbf{a}) \times \mathrm{constraints}(\mathbf{b})$.

- r_a and r_b are equivalent under rewriting based on m_n.
- $\mathrm{shared}(\mathbf{a},\mathbf{b})$ is the set of constraints included in both $\mathbf{a}$ and $\mathbf{b}$.
- $\mathrm{unshared}(\mathbf{a},\mathbf{b})$ = $\mathbf{a}$ - $\mathrm{shared}(\mathbf{a},\mathbf{b})$ and $\mathrm{unshared}(\mathbf{b},\mathbf{a})$ = $\mathbf{b}$ - $\mathrm{shared}(\mathbf{a},\mathbf{b})$ are the sets of constraints included in only one of the two constructions.
- $\mathrm{compare}_r(\mathrm{ctmap}(\mathbf{a},\mathbf{b},m_n))$ is the composite constituent type comparison, defined for total constituent maps as follows:

  $\approx$, if $\mathrm{unshared}(\mathbf{a},\mathbf{b}) = \mathrm{unshared}(\mathbf{b},\mathbf{a}) = \emptyset$

  $\prec$, if $|\,\mathrm{unshared}(\mathbf{a},\mathbf{b})| > 0$ and $|\,\mathrm{unshared}(\mathbf{b},\mathbf{a})| = 0$ ($\mathbf{a}$ is more constrained)

  $\succ$, if $|\,\mathrm{unshared}(\mathbf{b},\mathbf{a})| > 0$ and $|\,\mathrm{unshared}(\mathbf{a},\mathbf{b})| = 0$ ($\mathbf{b}$ is more constrained)

  $\sim$ otherwise, *i.e.*, if both $|\,\mathrm{unshared}(\mathbf{b},\mathbf{a})| > 0$ and $|\,\mathrm{unshared}(\mathbf{a},\mathbf{b})| > 0$

Figure 6.8. Constructional alignment definitions, including constituent maps, constraint maps and implicit type comparison functions.

relation, or in a sibling relation). This restriction is useful both for avoiding costly computation and for encouraging more conservative reorganization operations. Another restriction that improves efficiency is to limit the domain of search for valid constructional alignments to a subset of the full grammar, based (for example) on the most recently and frequently encountered constructions.

Example. A straightforward example of constructional alignment is illustrated below using our familiar THROW-BALL construction, along with a companion THROW-BLOCK construction (Figure 6.9). The best constituent alignment depends on existing constructional and schema type relations and distances; here we assume BALL and BLOCK are both subcases of a generic LABEL construction with form poles typed Word, and meaning poles (typed Ball and Block, respectively) both subcases of the Toy meaning schema. As shown below, the best constituent map m_n (*i.e.*, the one with minimal total type distances across domains) contains the predictable bindings between the same-typed THROW-BALL.t1 and THROW-BLOCK.x1 constituents, as well as between the THROW-BALL.t2 and THROW-BLOCK.x2 constituents (whose meanings are indirectly related). It is thus a total map whose composite comparison value $\text{compare}_n(m_n)$ is indirectly related ($\sim$). Using m_n to rewrite constraints in terms of the mapped constituents yields a total constraint map m_r, with two shared constraints. Since there are no unshared constraints, and all form and meaning poles are unspecified, the constraint comparison $\text{compare}_r(m_r)$ and the pole comparison compare_p are both implicitly equivalent ($\approx$), producing an overall constructional comparison of indirectly related ($\sim$).

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

construction THROW-BLOCK
  constituents
    x1 : THROW
    x2 : BLOCK
  form
    x1_f before x2_f
  meaning
    x1_m.throwee $\longleftrightarrow$ x2_m

Figure 6.9. Potential merge: The THROW-BALL and THROW-BLOCK constructions have compatible constituents and constraints.

(6–1) alignment(THROW-BALL,THROW-BLOCK):
m_n = { $\langle \text{t1} = \text{x1} \rangle$, $\langle \text{t2} \sim \text{x2} \rangle$ }
$\text{compare}_n(m_n) = \sim$
$\text{shared}(m_r)$ = { $\langle \text{x1}, \text{t1} \rangle_f$ before $\langle \text{x2}, \text{t2} \rangle_f$, $\langle \text{x1}, \text{t1} \rangle_m$.throwee $\longleftrightarrow$ $\langle \text{x2}, \text{t2} \rangle_m$ }
$\text{compare}_r(m_r) = \approx$
compare_p(THROW-BALL,THROW-BLOCK) = $\approx$
compare(THROW-BALL,THROW-BLOCK) = $\sim$

A similar exercise aligning THROW-BALL and the earlier-defined THROW-IMP yields the same final comparison ($\sim$), but with different values for the constituent comparison ($\prec$, based on the relation between Ball and Referent) and constraint comparison ($\succ$, based on the extra discourse-related constraint in THROW-IMP). In this case, the two constructions have global shared structure, but neither construction is strictly subsumed by the other.

6.3.2 Merging

To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes there were nothing but details, almost contiguous details.
— Jorge Luis Borges, "Funes, the Memorious"

The power of learning resides largely in the ability to extend one's reach beyond observed data: to grasp patterns only implicit in concrete experience and thereby avoid the perils of an overly replete grammar. In fact, we have already seen a limited form of generalization, in the discussion of the mapping heuristics from Section 6.2: any heuristic choosing a subset of bindings to map implicitly ignores some of the information available in the input, thereby in a sense forgetting many potential differences by not encoding them in the first place.

A more potent engine of generalization, however, is the merge operation, which exploits the constructional alignment information just defined to discern and extract common structure. Alignment not only suggests the best candidate pairs of constructions to merge; it also specifies precisely which categories and constraints are shared (and unshared). Specifically, two constructions are **mergeable** if they have global shared structure, and their domain poles and mapped constituents can be merged into more general constructional categories; unshared constraints can also be pruned away. Merging thus conflates two kinds of abstraction: type-based generalization of constituents and domain poles, and removal of non-overlapping constraints. Both of these allow merging to perform the same generalization function here as it does in other model merging applications, but with adaptations to allow it to handle constituent structure and relational constraints.

An algorithm for relational merging is given in Figure 6.10. The algorithm first aligns its two constructions, and then (if merging is possible) creates a new construction with the shared constituents and constraints, rewriting the original constructions as subcases of the newly merged construction retaining their unshared content. The main complicating consideration here is in step (2b) of the algorithm, in the merging of constituents. In the context of the language problem defined here, relatively little is assumed about the initial constructional categories; thus, depending on the stage of learning, the closest constructional ancestor $\mathrm{lca}(\mathbf{a}, \mathbf{b})$ may well be $\top$ or a relatively high-level, unconstrained category. Much richer structure is likely to be available in the form and meaning domains, providing a more concrete basis for positing a merged structure capturing the

Relational merge: Given relational constructions $\mathbf{a}$ and $\mathbf{b}$, return a merged construction $\mathbf{c}$ and rewritten constructions $\mathbf{a}'$ and $\mathbf{b}'$, or fail if no merge is possible.

1. Align $\mathbf{a}$ and $\mathbf{b}$, producing a constituent map $m_n = \mathrm{cmap}(\mathbf{a}, \mathbf{b})$ and constraint map $m_r = \mathrm{ctmap}(\mathbf{a}, \mathbf{b}, m_n)$. If $\mathrm{compare}(\mathbf{a}, \mathbf{b}) = \perp$, fail.
2. Create a new construction $\mathbf{c}$:
   (a) Set $\mathrm{type}(\mathbf{c}_f) = \mathrm{lca}_f(\mathbf{a}, \mathbf{b})$ and $\mathrm{type}(\mathbf{c}_m) = \mathrm{lca}_m(\mathbf{a}, \mathbf{b})$.
   (b) For each constituent pair $\langle n_a, n_b \rangle \in \mathrm{cmap}(\mathbf{a}, \mathbf{b})$,
   perform a *constituent merge* to produce the generalized type $\mathbf{g}_n = \mathrm{merge}_n(n_a, n_b)$ and add a constituent n of type $\mathbf{g_n}$ to $\mathrm{constituents}(\mathbf{c})$.
   (c) Add constraints in $\mathrm{shared}(\mathbf{a}, \mathbf{b})$ to $\mathrm{constraints}(\mathbf{c})$.
3. Rewrite $\mathbf{a}$ and $\mathbf{b}$ as subcases $\mathbf{a}'$ and $\mathbf{b}'$ of $\mathbf{c}$. For each item $\langle n_a, n_b \rangle \in \mathrm{cmap}(\mathbf{a}, \mathbf{b})$:
   - if $\mathrm{type}(n_a) < \mathrm{type}(\mathbf{c}.n)$, add the type constraint $n : \mathrm{type}(n_a)$ to $\mathbf{a}'$.
   - if $\mathrm{type}(n_b) < \mathrm{type}(\mathbf{c}.n)$, add the type constraint $n : \mathrm{type}(n_b)$ to $\mathbf{b}'$.

   Add constraints in $\mathrm{unshared}(\mathbf{a}, \mathbf{b})$ to $\mathbf{a}'$, and the constraints in $\mathrm{unshared}(\mathbf{a}, \mathbf{b})$ to $\mathbf{b}'$.

Figure 6.10. Relational merge algorithm

commonalities between constituents. The search for a merged constituent type can avail itself of structure in any of the three domains by hypothesizing new constructional types that capture this shared structure. In effect, generalization at the relational level can induce generalization at the constituent level, thereby potentially effecting the creation of new constructional categories.

Constituent merge: Given constructional constituents of type $\mathbf{a}$ and $\mathbf{b}$, return their generalized construction type, adding a new constructional category if appropriate.

1. Let $\mathbf{g} = \mathrm{lca}(\mathbf{a}, \mathbf{b})$ (*i.e.*, the closest common constructional ancestor of $\mathbf{a}$ and $\mathbf{b}$).
2. If $(\mathrm{distance}(\mathrm{type}(\mathbf{g}_m), \mathrm{lca}_m(\mathbf{a}, \mathbf{b})) \leq \mathrm{distance}_{max})$ and
   $(\mathrm{distance}(\mathrm{type}(\mathbf{g}_f), \mathrm{lca}_f(\mathbf{a}, \mathbf{b})) \leq \mathrm{distance}_{max})$
   return $\mathbf{g}$.
3. If $(\mathrm{distance}_m(\mathbf{a}, \mathbf{b}) \leq \mathrm{distance}_{max})$ and $(\mathrm{distance}_f(\mathbf{a}, \mathbf{b}) \leq \mathrm{distance}_{max})$:
   - Create $\mathbf{g}'$ with $\mathrm{type}(\mathbf{g}'_f) = \mathrm{lca}_f(\mathbf{a}, \mathbf{b})$ and $\mathrm{type}(\mathbf{g}'_m) = \mathrm{lca}_m(\mathbf{a}, \mathbf{b})$.
   - Add $\mathbf{g}$ to $\mathrm{parents}(\mathbf{g}')$, and add $\mathbf{g}'$ to $\mathrm{parents}(\mathbf{a})$ and $\mathrm{parents}(\mathbf{b})$.

   else fail.
4. For all constructions $\mathbf{x}$:
   If $\mathbf{x} \prec \mathbf{g}'$ and the *inductive leap condition* $\mathrm{ilc}(\mathbf{x}, \mathbf{g}')$ is satisfied,
   add $\mathbf{g}'$ to $\mathrm{parents}(\mathbf{x})$.
5. Return $\mathbf{g}'$.

Figure 6.11. Constituent merge algorithm

The constituent merging algorithm in Figure 6.11 first checks whether the current closest constructional ancestor will suffice, *i.e.*, whether its form and meaning types are within an upper bound

$\mathrm{distance}_{max}$ of the respective type generalizations of the constituents being merged. If not, it creates a new, more specific category whose form and meaning poles are typed appropriately; this category $\mathbf{g}'$ is a subtype of the (erstwhile) constructional generalization $\mathbf{g}$, and a new parent of the types of the two merged constituents.

At this point, the hypothesized constructional type has only two explicit subtypes. The larger hypothesized relational construction thus does not yet extend to any territory beyond the existing (relational) constructions. For that, additional generalization is required. Step 4 of the algorithm in Figure 6.11 effects this generalization by seeking other candidate constructions to subsume; an appropriate candidate is defined here as one that is (implicitly, *i.e.*, by virtue of its pole types) more specific than the new constructional type $\mathbf{g}'$,[10] and further that satisfies an **inductive leap condition**. Intuitively, this condition controls how permissively or restrictively the model sanctions extensions of the new constructional category on the basis of its current subcases; it should thus depend on factors such the similarity (*e.g.*, distance in the type hierarchy) between $\mathbf{x}$ and the current subcases, or the degree of overlap between the form and meaning types of $\mathbf{x}$ and $\mathbf{g}$.

An extremely permissive inductive leap condition, for example, would be to automatically extend $\mathbf{g}'$ to include all implicitly subsumed categories. In other words, since the pole types of $\mathbf{g}'$ are defined as generalizing those of $\mathbf{a}_m$ and $\mathbf{b}_m$, any construction whose pole types are subsumed by those of $\mathbf{g}'$ would also be made an explicit constructional subcase of $\mathbf{g}'$. A more restrictive condition may require the explicit inclusion of many or most of a construction's closest relatives before permitting an inductive leap (perhaps better characterized in this case as an inductive hop). The particular inductive heuristic employed by the current model focuses on the meaning domain, permitting leaps to include $\mathbf{x}$ if more than half, or at least three, of its semantic descendents (that is, subcases of its meaning type $\mathrm{type}(\mathbf{x}_m)$) are among the meanings of the explicit subcases of $\mathbf{g}'$. This corresponds roughly to the notion that observing either half the members of a class, or at least three members of a class, in a particular constituent role is sufficient evidence on which to infer that other members of that class are also likely to be allowed in that constituent role.

Example merge operations

Recall the constructional alignment example from Section 6.3.1, based on the THROW-BALL and THROW-BLOCK constructions in Figure 6.9. As noted earlier, the two constructions have global shared structure, and based on their constituent and constraint maps they are implicitly indirectly

[10] These implicit subtypes can be found by aligning $\mathbf{g}'$ with all existing constructions. Note that this process may determine that some constructions are (implicitly) equivalent; equivalent constructions are folded together into a single category whose subcases are the union of the combined categories' subcases.

related. They are thus a good candidate for merging, using the relational and constituent merging procedures: a new construction with the same number of constituents can be hypothesized, each corresponding to a constituent pair in the constituent map.

In this case, the first constituent to add, corresponding to the constituent pair $\langle$x1, t1$\rangle$, poses no difficulties: they are both typed as the construction THROW, so their generalized type is simply THROW, which is returned by the constituent merging algorithm as the type for the merged constituent. The second constituent, based on the constituent pair $\langle$x2, t2$\rangle$, is more complicated. An attempted merge between these two might have several different outcomes, depending on the state of current constructional and schema knowledge. Below we consider two conditions.

Generalization on the construction domain. Let us assume first that the BALL and BLOCK constructions have an existing parent, a TOY constructional category that consists of words used to denote toys. Assume also that their associated meaning schemas, Ball and Block, are meaning siblings, sharing a parent Toy schema. This situation is depicted in the top portion of Figure 6.12. Each of the rectangles in the central area corresponds to an existing construction, with constructional constituents iconically represented as nested constructional rectangles. The two relational constructions to be merged are the upper and lower boxes, each connected to information in the form domain (left) and the meaning domain (right).[11] The overlapping shaded portions of the form and meaning domains are intended to convey information shared by the two constructions, such as the presence of particular words and meanings, both concrete and more abstract.

Selected type inheritance links, including the above-noted relations, are shown as vertical lines starting at a construction or schema category and ending in a branched pointer to its subsumed category. In particular, the highlighted link between the constituent BALL and BLOCK constructions and their parent TOY construction serves as the basis for generalization. That is, since the existing parent type TOY has form and meaning types that are as specific as the generalized form and meaning types for the BALL and BLOCK constructions, the constituent merging algorithm proceeds as for the first constituent.

The resulting merged construction is depicted in the lower portion of Figure 6.12, with heavy dashed lines around the new relational construction, as well as its inheritance links to the two (rewritten) subcases. This new THROW-TOY construction is defined in ECG notation in Figure 6.13. The constituents are named according to their THROW-BALL analogue, as are the shared constraints. The original constructions are rewritten as subcases to omit the shared (inherited) content, but each includes a specializing type constraint on the (merged) t2 constituent.

[11] For clarity, this depiction omits the structured constituent mappings to the form and meaning domains.

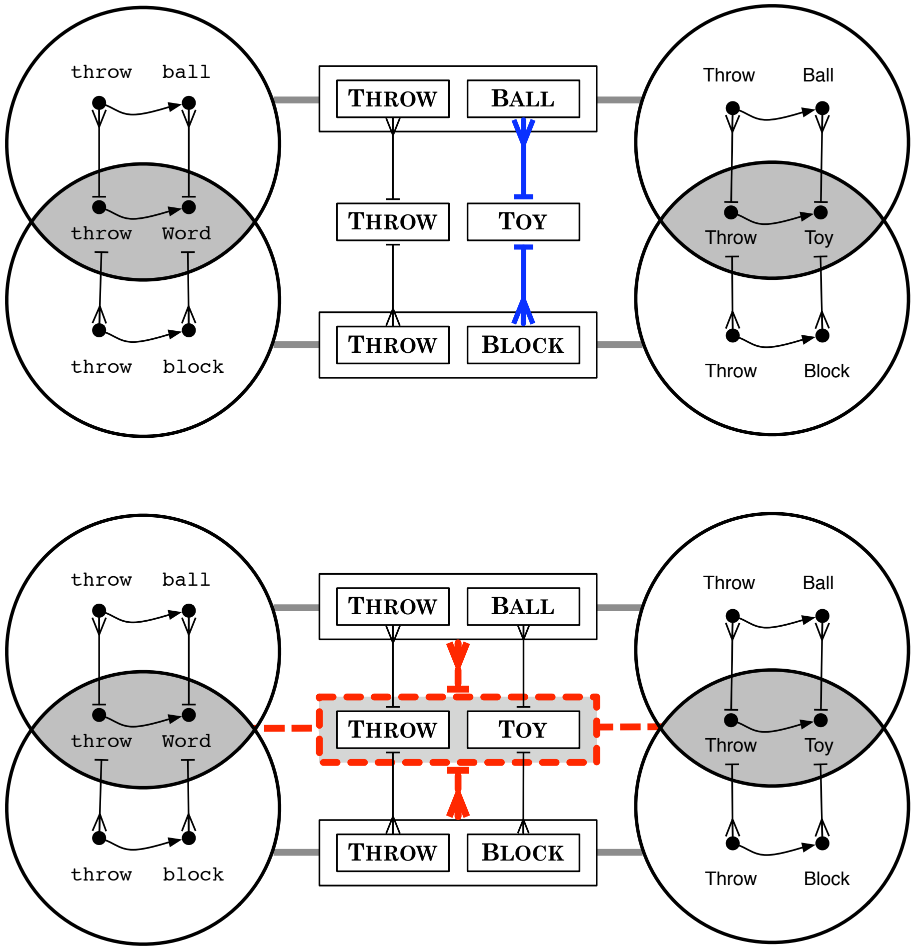

Figure 6.12. Construction-based generalization: Merging THROW-BALL and THROW-BLOCK, assuming BALL and BLOCK have a common parent, the TOY construction (where inheritance relations are indicated by the vertical lines from a categories branching to subcases).

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

construction THROW-BALL
  subcase of THROW-TOY
  constituents
    t2 : BALL

construction THROW-BLOCK
  subcase of THROW-TOY
  constituents
    t2 : BLOCK

Figure 6.13. Result of merge operation: A new THROW-TOY construction contains the shared content of the merged constructions, while the rewritten THROW-BALL and THROW-BLOCK constructions specialize type of the merged constituent.

Generalization and category formation on the meaning domain. Now consider an alternate situation in which no common constructional parent is available for the mapped constituents. As shown in Figure 6.14, Ball and Block have a common ancestor (the Label construction), but it is not an immediate parent, and its meaning type is not as specific as the generalization available for the constituent pair ⟨x2, t2⟩ (as before, a Toy category). The constituent merging algorithm is thus prompted to posit a new (lexical) constructional category, based on the nearest common ancestors in the form and (most relevantly) meaning domains. That is, it is similarity (*i.e.*, type proximity, via a shared parent) in the meaning domain (shown as a heavy inheritance link) that drives the constituent merge.

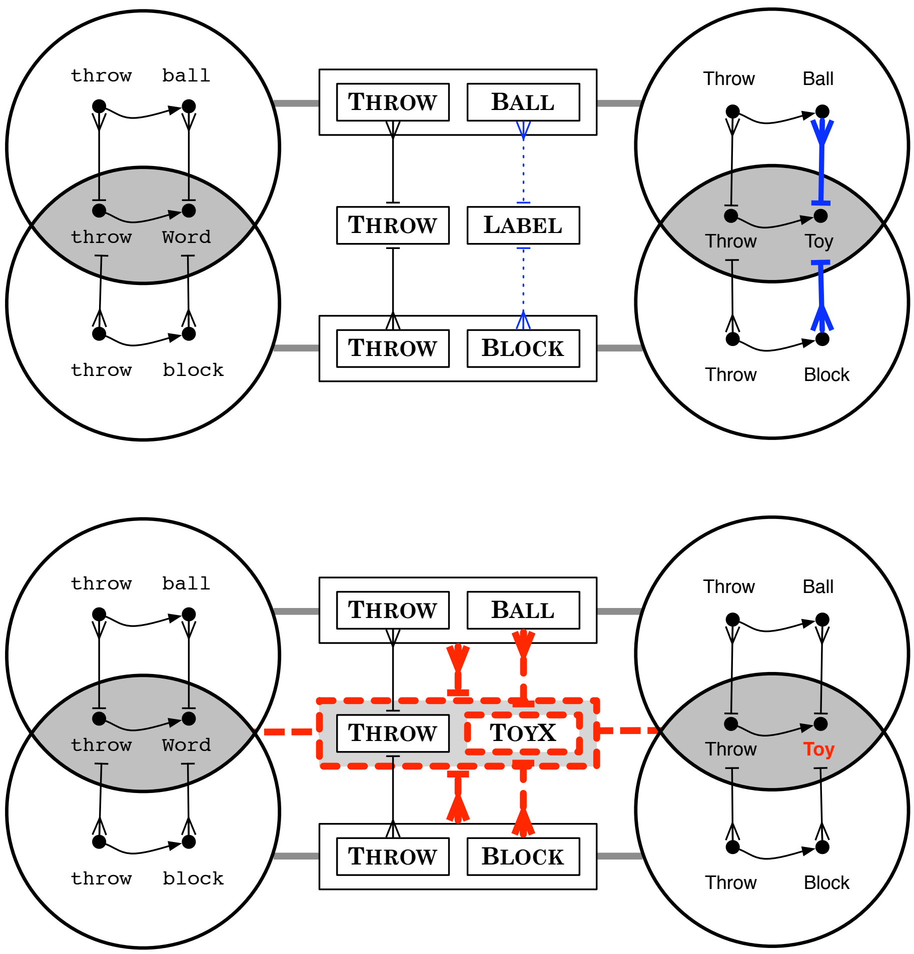

Figure 6.14. Meaning-based generalization: Merging Throw-Ball and Throw-Block, assuming Ball and Block have no common constructional parent (where the dotted line indicates a non-immediate subcase relationship to a general Label category). But a constituent merge is licensed by their common meaning parent Toy, resulting in a new Throw-ToyX relational construction and a new ToyX lexical construction.

The resulting ToyX constructional category is used as the type constraint on the merged constituent of the relationally merged construction, as depicted in the lower portion of Figure 6.14; heavy dashed lines mark both the learned constructions and the new constructional inheritance links to their respective constructional subcases. In effect, the ToyX construction *specializes* the general Label constructional category to one that captures the shared meaning of the two constructions. Figure 6.15 shows the result of this alternative merge scenario in ECG notation, with a common schematic category of Toy licensing the formation of the new constructional ToyX category. As suggested earlier, however, the ToyX category (and, by extension, the Throw-ToyX construction) may not immediately expand the learner's linguistic repertoire. As defined, only the Ball and Block are valid subcases; the inductive leap has not yet occurred. Depending on the restrictiveness of the inductive leap condition, semantically related constructions (*i.e.*, other constructions whose meaning poles are typed as Toy) may, sooner or later, be allowed into the ToyX category; once they are, the Throw-ToyX construction can license previously unencountered utterances involving the throwing of any kind of toy.

construction Throw-ToyX
  constituents
    t1 : Throw
    t2 : ToyX
  form
    $t1_f$ before $t2_f$
  meaning
    $t1_m$.throwee $\longleftrightarrow$ $t2_m$

construction ToyX
  subcase of Label
  form : Word
  meaning : Toy

construction Ball
  subcase of ToyX
  meaning : Ball

construction Block
  subcase of ToyX
  meaning : Block

Figure 6.15. Constructions resulting from meaning-based merge, showing new constructional category ToyX specializing the meaning type of the Label category. (The two rewritten original constructions are similar to those in Figure 6.13 and not shown here.)

Merging adaptations and heuristics

The merge operation defined here departs from previous versions of model merging in several ways. First, the two original constructions are not removed from the grammar but instead are retained and rewritten as subcases of the new, more general construction. This approach is motivated by the assumption that linguistic knowledge of differing degrees of specificity can coexist within the grammar, and there is no evidence that children lose constraints that may apply to more specific constructions once they learn more general constructions. Of course, over time, the more general constructions are likely to prove more widely applicable and therefore be reinforced over time. Under certain circumstances, the dominance of the merged construction may be such that the specific constructions are rarely used in practice and therefore eventually drop to such a low frequency that they are effectively pruned. Alternatively, if the specific constructions are frequent or contribute information not available from the general case, they may be worth keeping in the grammar, even at the expense of having to compete with the general construction.[12] In either case, the retention or elimination of specific constructions hinges on the long-term dynamics of use and is not an automatic consequence dictated by the merge generalization.

Other changes are motivated by the multi-dimensional type system of the construction domain. The basic notion of generalizing to a shared parent type, or positing a new one, is analogous to the creation of new non-terminal symbols and their rule expansions in domains like the context-free grammars and probabilistic attribute grammars addressed by Stolcke (1994). In the construction grammar domain, we modify these to put form and meaning on equal footing as potential sources of generalization, and allow explicit inheritance relations to be acquired. The resulting category structure, while not able to capture the full range of phenomena associated with radial categories (Chapter 2), nevertheless has the basic property of flexibly associating categories of form with categories of meaning. It is also noteworthy that the formation of new categories over constituent constructions is prompted by the relational merge operation. This operation endows the model with the capacity to learn constructional categories that can, in principle, be motivated by conceptual (meaning) categories as well as the syntactic context in which a category appears, thus providing for aspects of both semantic and syntactic bootstrapping (as discussed further in Chapter 9).

In general, the space of possible merges can grow very large. As noted earlier, the construction alignment procedure could exhaustively compare every construction pair, checking every constituent and constraint map; the merging algorithm could include any subset of shared constraints in the new construction, and permit constituent merges based on varying type distances (*i.e.*, more or less closely related types). To keep the space manageable and tie the merging process more closely to the incoming flow of data, the model exploits several simplifying assumptions:[13]

- Constituent merges are conservative, with $\mathrm{distance}_{max}$ set to 1. That is, only shared immediate parents sanction either constructional generalization (with no category formation) or meaning-based generalization (with category formation).

- Candidate merges are restricted to those used within a small recency window; these corre-

[12]Such a process is undoubtedly also useful for the acquisition of idiomatic constructions.

[13]Further control over triggering conditions may be possible with more integration with the analyzer. As mentioned in Section 4.2.1, the analysis process builds a chart of possible sets of matching constructs. If the analyzer records near-miss constructions (*i.e.*, those whose constraints come within a small window of being satisfied), then input tokens that do not yield a spanning analysis may suggest possible near-miss constructions as merge candidates.

spond to primed constructions in short-term memory that serve as more accessible candidates for comparison and generalization.

- Candidate merges are restricted to those with a relative frequency above a minimum threshold, to prevent needless merging based on low-usage constructions.
- Redundant merges are avoided by collapsing proposed categories that are implicitly type-equivalent to an existing category.

Other ways of circumscribing the merge operator may also be desirable, in view of developmental evidence that generalizations apply earlier and more consistently for entities and locations than for action words. Such preferences may be modeled as differing maximum neighborhoods for different portions of the construction and schema hierarchies.

6.3.3 Joining

The remaining two operations are relatively simple, manipulating overlapping constituent structure as a form of learning. The join operation creates a new construction that is the union of the constituents and constraints of two existing constructions. The algorithm below merges the (mapped) constituents and constraints that are part of both constructions and adds the remaining constituents and constraints to the result, all with the appropriate rewrites.

Join: Given constructions $\mathbf{a}$ and $\mathbf{b}$, return a joined construction $\mathrm{join}(\mathbf{a},\mathbf{b})$.

1. Align $\mathbf{a}$ and $\mathbf{b}$, resulting in a constituent map $\mathrm{cmap}(\mathbf{a},\mathbf{b})$ and shared constraints $\mathrm{shared}(\mathbf{a},\mathbf{b})$. For each $\langle n_a, n_b\rangle$ in $\mathrm{cmap}(\mathbf{a},\mathbf{b})$, require $\mathrm{type}(n_a)$ = $\mathrm{type}(n_b)$.
2. Create a new construction $\mathbf{c}$, with:
   - a constituent n of type $\mathrm{type}(n_a) = \mathrm{type}(n_b)$ for each item $\langle n_a, n_b\rangle$ in $\mathrm{cmap}(\mathbf{a},\mathbf{b})$;
   - any constituents of $\mathbf{a}$ and $\mathbf{b}$ not mapped by any pair in $\mathrm{cmap}(\mathbf{a},\mathbf{b})$;
   - constraints in $\mathrm{shared}(\mathbf{a},\mathbf{b})$;
   - constraints in $\mathrm{unshared}(\mathbf{a},\mathbf{b})$; and
   - constraints in $\mathrm{unshared}(\mathbf{b},\mathbf{a})$,

   where constraints are rewritten according to $\mathrm{cmap}(\mathbf{a},\mathbf{b})$ and required to be compatible.

For example, the general HUMAN-THROW and THROW-BOTTLE constructions shown in Figure 6.16 satisfy the conditions of the join algorithm, with a shared constituent of type THROW and one unshared constituent from each construction (*i.e.*, the constituent map $m_n = \{\langle \mathsf{t} = \mathsf{t1}\rangle\}$). While none of the constraints are shared, they are compatible in both form and meaning (in the latter case, based on filling different roles of the same Throw schema). The result of applying the join operation is shown in Figure 6.17.

construction HUMAN-THROW
constituents
h : HUMAN
t : THROW
form
h_f `before` t_f
meaning
t_m.thrower $\longleftrightarrow$ h

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

Figure 6.16. Potential constructions for joining: the HUMAN-THROW and THROW-BOTTLE constructions have a common compatible constituent.

construction HUMAN-THROW-BOTTLE
constituents
h : HUMAN
t : THROW
p : BOTTLE
form
h_f `before` t_f
t_f `before` p_f
meaning
t_m.thrower $\longleftrightarrow$ h_m
t_m.throwee $\longleftrightarrow$ p_m

Figure 6.17. Result of join operation: new construction HUMAN-THROW-BOTTLE includes constituents and constraints of the two joined constructions.

Like the other reorganization operations, candidate joins can be found via constructional alignment and a search over known constructions. In practice, however, this is far too permissive a criterion, since so many constructions overlap in some part. Instead, the model applies are more specific triggering condition, based on the attested co-occurrence of two constructs during analysis that are competing for the same constituent and that otherwise have overall compatible constraints.

Figure 6.18 depicts this situation, focusing on the constructional domain only. In the example, analyzing the sentence *You throw the bottle* (ignoring the article) might result in two partial analyses, each involving one of the constructions in Figure 6.16 and spanning two words (*You throw* and *throw bottle*). These can be seen as competing for the THROW-typed constituent. Since the relevant form and meaning constraints can all be satisfied by unifying this constituent, this chart-based co-occurrence can lead to the formation of the joined construction including all three constituents. (See Chapter 4 for details about the role of the chart in analysis.)

Notice that the join operation is in some ways quite similar to a mapping operation, since it creates larger constructions out of already available constituent chunks. (Alternatively, a relational map might be seen as joining two constructions in the chart with *no* overlap.) The key difference is that while the join operation may create relational constructions by dint of unifying the relational

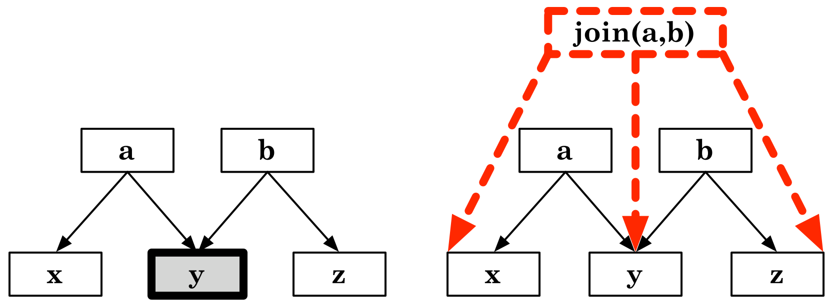

Figure 6.18. Joining two constructions with compatible constituents

constraints of the original constructions (as in the example above), it does not in itself identify any new relational constraints of the kind produced by relational mapping.

6.3.4 Splitting

The split operation is the inverse of the join operation: both exploit the presence of constructional overlap, but splitting creates new constructions based on non-overlapping, or unshared, subsets of existing constructions. It is thus the conceptual correlate of taking the set difference as the basis for learning, defined in the split algorithm below analogously to the join operation. As shown here, it is asymmetric, hypothesizing a construction that includes just those constituents and constraints that are part of $\mathbf{a}$ but not $\mathbf{b}$.

Split: Given constructions $\mathbf{a}$ and $\mathbf{b}$, return a split construction $\mathrm{split}(\mathbf{a}, \mathbf{b})$.

1. Align $\mathbf{a}$ and $\mathbf{b}$, resulting in a constituent map $\mathrm{cmap}(\mathbf{a}, \mathbf{b})$ and shared constraints $\mathrm{shared}(\mathbf{a}, \mathbf{b})$.
2. Create a new construction $\mathbf{c}$, containing:
   - any constituents of $\mathbf{a}$ not mapped by any pair in $\mathrm{cmap}(\mathbf{a}, \mathbf{b})$, and
   - all constraints in $\mathrm{unshared}(\mathbf{a}, \mathbf{b})$.

As with the join operation, the space of possible splits makes it impractical to search exhaustively based on partially overlapping constructions; again, it is the coincident appearance of such overlapping constructions in the course of experience that triggers a split operation. Our current model restricts splitting for the special case of learning meanings for unknown forms within a larger relational construction, where the larger construction may correspond closely to an entire input token.[14] The presence (*i.e.*, in the chart of identified constructs) of another construction that

[14]Such a construction could result, for example, from using the "map all" heuristic in Section 6.2.

accounts for a subset of the larger spanning construction can be treated as grounds for splitting the spanning construction. Cast in the terms above, **a** is the spanning construction; **b** is informationally a subset of **a**; and the remainder is a form in $\mathbf{a}_f$ that has no stable lexical mapping. In this case, **b** may seen as *explaining away* part of the the content of **a**, thus focusing the learner's attention on the remainder. The situation is depicted graphically in Figure 6.19.

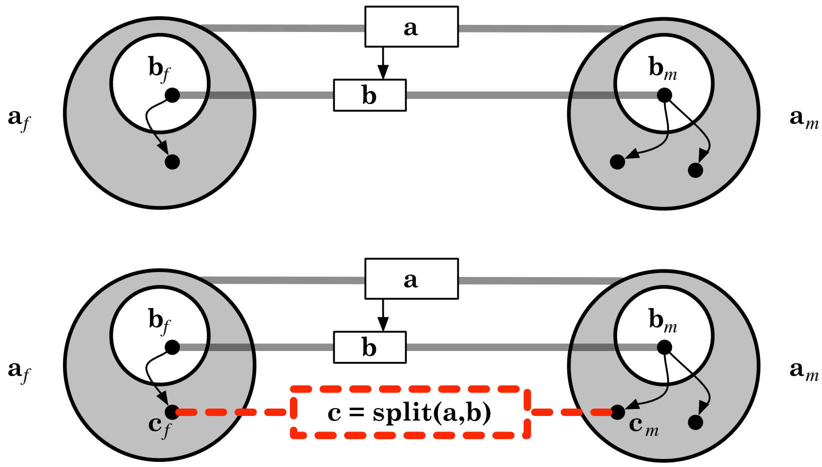

Figure 6.19. Splitting a construction based on unshared content: The split operation maps elements in **a** and not shared by **b**.

A pair of constructions potentially leading to a split is shown in Figure 6.20. The THROW-FRISBEE construction is defined as having an unknown form with orthography "frisbee"; this form does not yet have a learned lexical meaning; the Frisbee in its meaning pole is a meaning constraint on the thrown object in the scene, and not yet directly associated with the "frisbee" form.[15] But the known THROW construction's content (and any constraints defined on that content) can in effect be subtracted from the THROW-FRISBEE construction, allowing a new FRISBEE construction to be inferred based on the remaining unknown word and meaning schema. Once learned, the new construction can be used to rewrite the THROW-FRISBEE construction, as shown in Figure 6.21, resulting in a more concise (and structurally isomorphic) relational construction.

Besides facilitating simple word learning, the splitting operation could be applied to learn segments of any granularity. In particular, it may be useful for learning morphological markers, given some mechanisms for noticing word-internal structure. The current work does not address the complexities of morphology learning; however, the same principles of relational structure apply to that problem, so all of the operations proposed here have direct word-internal analogues.

[15]This assumes that unfamiliar words are segmentable and identified as such when they appear in larger constructions.

construction THROW-FRISBEE
constituents
t1 : THROW
t2 : UNKNOWN
form
$t1_f$ `before` $t2_f$
$t2_f$.orth $\longleftarrow$ “frisbee”
meaning
$t1_m$.throwee : Frisbee

construction THROW
form
$\textbf{self}_f$.orth $\longleftarrow$ “throw”
meaning
$\textbf{self}_m$: Throw

Figure 6.20. Potential constructions allowing a split operation

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

construction FRISBEE
form
$\textbf{self}_f$.orth $\longleftarrow$ “frisbee”
meaning
$\textbf{self}_m$: Frisbee

Figure 6.21. Result of split operation: the THROW-FRISBEE construction is rewritten using the newly learned FRISBEE construction.

6.4 Summary

Our tour through the model’s learning operations is now complete. This chapter has presented a number of strategies for constructing (and deconstructing) linguistic knowledge based on experience. The proposed operations can be conceptualized as inferring new structure from context (through mapping operations) and rearranging existing structures into new combinations (through the reorganization operations), but it should be clear from the discussion that each operation has a wide range of interpretations, and there may be several overlapping routes to similar grammars.

A variety of heuristics have been presented for zeroing in on the most auspicious conditions for applying each of the operations; these are summarized below and depicted in Figure 6.22:

- Input tokens containing unknown forms may be **split** into component pieces corresponding to an already known construction and a newly learned construction for the unknown form.

- After analysis, any overlapping analyses may be **joined** based on any shared constituent for which they may be competing.

- After resolution, any contextually available bindings not explained by linguistic analysis may be **mapped** into a new form-meaning pair.

- Known constructions can periodically be **merged** to yield new constructional categories capturing their shared structures.

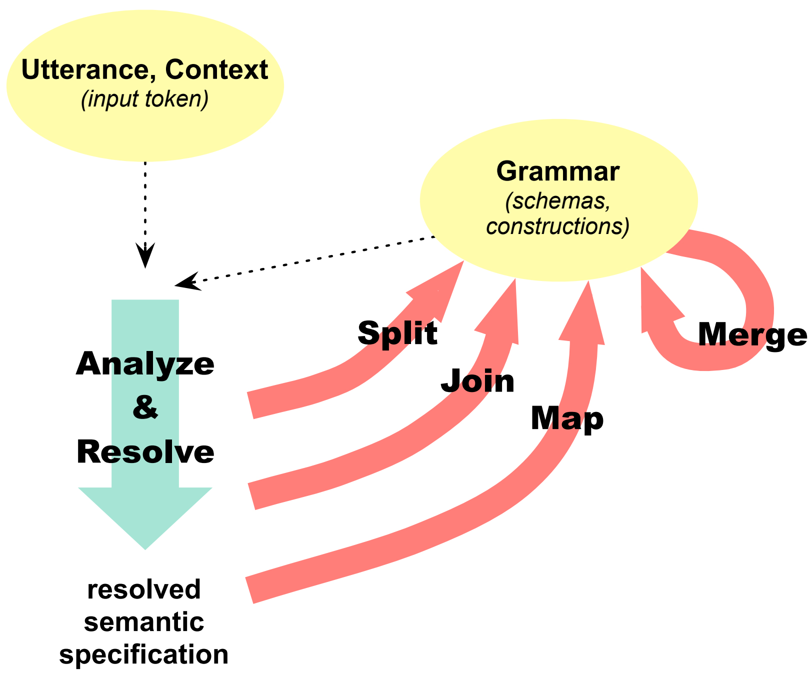

Figure 6.22. Overview of usage-based learning operations: the map, merge, join and split operations update the grammar, drawing on the results of analyzing and resolving the utterance in context.

A few caveats about the proposed learning algorithms are worth mentioning. Clearly, the toy examples we have described have been maximally simplified for expository purposes. In fact, the number of conceptual frames and entities in a given situation representation is likely to be much larger and give rise to many more potential candidate operations. Thus, when the algorithm has several possible operations that result in comparable improvements to the grammar, a few unlucky choices may result in a considerable setback in learning.

It is difficult to provide direct evidence for the precise learning strategies proposed. The toolkit of operations available to a child learner might be much larger, or it may require variants of the operations proposed here that are more finely attuned to specific triggering circumstances. Moreover, any attempt to address the full array of cues involved in, for example, the acquisition of morphological paradigms, would clearly require these operations to be augmented with (at least) principles like those proposed by Peters (1985) for segmentation purposes.

Nonetheless, the proposed operations are well-motivated on both theoretical and empirical grounds. From a representational perspective, these operations address the key challenge of first acquiring and then reorganizing structured relational mappings. They are designed so that all parts of the hypothesis space are theoretically within the learner's reach. Thus, assuming the learner continues to receive helpful data indefinitely — as seems to be the case for most child learners — valid correlations should occur frequently and reliably enough to overwhelm bad choices, and the potential pitfalls of wandering into impoverished neighborhoods in the space of grammars may be relatively minimal.[16]

Evidence for some aspects of each operation can also be found in the cognitive and developmental literature. As mentioned earlier, the structures and processes of the model are consistent with a large body of work in usage-based theories of language, as well as some general principles of categorization (Lakoff 1987). The most obvious parallels are with Slobin's (1985) Operating Principles for learning grammar, which include operations for similarity-based merging and segmentation; preferences for mapping onto identifiable meanings and relations relevant to adjacent referential units; and preferences for finding meanings for particular word orders. More recent work suggests that much of children's early syntactic creativity can be accounted for as relatively minor adjustments to previous utterances (Tomasello 2003; Lieven *et al.* 2003). Interestingly, Lieven *et al.* (2003) identify a number of operations for changing a source (observed) utterance to a target (desired) utterance, including the substitution of a morpheme or word in a slot-filler position and the addition of a morpheme or word to a previous utterance. Both of these operations have clear correlates in the operations proposed here.

Thus, while these operations, like other aspects of the model, must be further refined to account more fully for statistical or developmental phenomena, they provide at least a first cut at a linguistically and developmentally motivated set of heuristics for navigating through the space of grammars. The task of assessing which of these operations lead to the most promising regions of this space is entrusted to the model's evaluation criteria, to which we turn next.

[16]Indeed, the contextual mapping operations might be seen as broadly inspired by a Gricean 1975 attitude toward one's interlocutors: the operations are only useful insofar as the input utterances accurately, succinctly and clearly capture a relevant communicative intention.

Chapter 7

Simplicity-based evaluation heuristics

It is superstitious to put one's hopes in formalities,
but arrogant to refuse to submit to them.
— *Blaise Pascal*

The last chapter established several cognitively motivated learning operations for updating the set of constructions. Each of these operations is triggered by a particular instance of language use — *i.e.*, it is motivated by the results of applying the analysis process described in Chapter 5 to a given input token. Accordingly, any proposed update is presumed to be helpful in at least one specific scenario. But how can the learner be sure that a given operation will be beneficial on a wider and more lasting scale? Likewise, since multiple operations may be proposed based on a single token, how should the learner choose among several options?

We rely on information theory for guidance in these matters. Section 7.1 reviews the minimum description length principle (Rissanen 1978) and its relation to Bayesian inference. Section 7.2 defines MDL metrics for our model space, ECG constructions. Section 7.3 then revisits the operations illustrated in the last chapter and shows their effects on the proposed MDL length functions. Some refinements to the length criteria are considered in Section 7.4, followed by discussion in Section 7.5.

7.1 Minimum description length

The task of choosing the best grammar (or best update to the current grammar) is a special case of the pervasive problem of **model selection**, where *model* takes its most general reading here as a putative account for or hypothesis about some observed data. As reviewed in Section 5.3.2, the Bayesian approach is to choose the model M that maximizes the *posterior* probability $P(M|D)$ given the observed data D:

$$\hat{M} = \underset{M}{\text{argmax}} \; P(M|D) \tag{7.1.1}$$

$$= \underset{M}{\text{argmax}} \; \frac{P(M) \cdot P(D|M)}{P(D)} \tag{7.1.2}$$

$$= \underset{M}{\text{argmax}} \; P(M) \cdot P(D|M) \tag{7.1.3}$$

$P(M)$ and $P(D)$ are *prior* probabilities for the model and data, respectively, and $P(D|M)$ is the *likelihood* of the model having generated the data D. Since D is fixed for a particular task, we can ignore it for the purposes of maximizing $P(M|D)$, resulting in (7.1.3).

Applications of Bayesian inference have become the *de facto* coin of the machine learning realm. Both terms in (7.1.3) are often convenient to obtain: the likelihood $P(D|M)$ is usually defined by the model, while the prior $P(M)$ can be defined by the modeler to reflect domain-specific biases. The choice of best model can thus hinge on both prior expectations and actual observations.

While this purely probabilistic formulation lends itself to many domains, it is not always feasible to construe the data as being generated from a probability distribution, nor to construe the prior as a degree of belief about distributions. Some domains may be better suited to a related *information-theoretic* approach to model selection, in terms of **description lengths**. The motivating intuition derives from William of Occam's trusty razor, the principle that entities should not be multiplied beyond necessity — and that, all else being equal, simpler hypotheses are to be preferred. In this context, the hypothesis can be thought of as a *code* for representing data, and potentially compressing it by exploiting its regularities. The *simplest* hypothesis is thus the code that permits the most succinct description of the data, or its most *compact* encoding.

Viewing the task as a search for a message's optimal code length brings us back, conveniently, to probabilities, and the one-to-one correspondence between probability distributions and code length functions (established by the Kraft-McMillan inequality; see MacKay (2003) for an introduction). In particular, for any code C, there is a corresponding probability distribution such that the length $\mathcal{L}(x)$ of a message x encoded using C is equal to its inverse log probability, and vice versa:

$$\mathcal{L}(x) = \log_2 \frac{1}{p(x)} \tag{7.1.4}$$

That is, the optimal code is one that assigns the shortest codes to the most probable messages, thus achieving the maximal possible *compression* of information.

Returning to model selection, we can apply this principle — that the optimal coding length of message is equal to its inverse log probability (Li & Vitányi 1997) — to measure both the description length (or *cost*) $\mathcal{L}(M)$ of a particular model M and the data length $\mathcal{L}(D|M)$ of the observed data D as encoded (or predicted) by M. The one that minimizes this sum, *i.e.*, the one with **minimum description length** (Rissanen 1989), is chosen as the best model $\hat{M}$:

$$\begin{aligned}\hat{M} &= \underset{M}{\text{argmin}} \;\; \mathcal{L}(M) + \mathcal{L}(D|M) \\ &= \underset{M}{\text{argmin}} \;\; -\log P(M) - \log P(D|M)\end{aligned}$$

This sum is related to the posterior probability from (7.1.1)–(7.1.3) above:

$$\begin{aligned}\hat{M} &= \underset{M}{\text{argmax}} \;\; P(M|D) \\ &= \underset{M}{\text{argmax}} \;\; \log(P(M) \cdot P(D|M)) \\ &= \underset{M}{\text{argmax}} \;\; \log P(M) + \log P(D|M) \\ &= \underset{M}{\text{argmin}} \;\; -\log P(M) - \log P(D|M) \\ &= \underset{M}{\text{argmin}} \;\; \mathcal{L}(M) + \mathcal{L}(D|M)\end{aligned}$$

While the Bayesian and information-theoretic views differ somewhat in underlying motivation, they afford many of the same advantages when applied to model selection. Both allow the choice of model to depend on the tradeoff between inherent aspects of the model and its closeness of fit to the data — where the Bayesian interpretation favors models that are *a priori* more likely than others, while the MDL interpretation is noncommittal about such prior distributions and evaluates models purely in terms of the information required to describe them. These views are in harmony when the prior expectations explicitly favor model simplicity: richer, more structured and more informative models incur greater MDL costs and are *a priori* less likely than simpler models, but these costs may be offset by the correspondingly simpler analyses of the data they may allow. I will exploit both formulations in the discussion to follow.

7.2 MDL for ECG

The duality between description length and probability provides a useful framework for our current task. Assuming ECG grammars as the model space, the optimal grammar $\hat{G}$ given the data

(7.2.1) is the one with the shortest **total description length** (7.2.2):

$$\begin{aligned}\hat{G} &= \underset{G}{\operatorname{argmax}}\ P(G|D) && (7.2.1)\\ &= \underset{G}{\operatorname{argmax}}\ P(G) + P(D|G) \\ &= \underset{G}{\operatorname{argmin}}\ \mathcal{L}(G) + \mathcal{L}(D|G) && (7.2.2)\end{aligned}$$

We will refer to the first term in (7.2.2) $\mathcal{L}(G)$ as the **grammar length** (the number of bits required to encode the grammar) and the second term $\mathcal{L}(D|G)$ as the **data length** (the number of bits required to encode the observed data using the grammar).

In general, viewing grammars in terms of description length is more convenient for our model space, particularly for the grammar length. ECG grammars as defined in Chapter 3 are collections of discrete, declarative structures, over which a prior distribution is at best difficult to compute. The data length may be somewhat more amenable to a probabilistic view: the constructional scoring metric given in Section 4.3.1 approximates the conditional probability of an analysis given a grammar, though the token score (including metrics for form and meaning) is not probabilistic.[1] It is thus easier in this situation to cast grammar learning as an encoding problem. Informally:

- The grammar length is the sum of the lengths of its constructions. The length of a construction is defined in terms of length metrics over all of its associated information, such as its constituents, type constraints, form and meaning constraints and constructional relations.
- The data length is the sum of the lengths of the input tokens in the corpus. The length of an input token is defined in terms of the constructs involved in the best analysis, as well as any additional the cost of encoding any discrepencies between the analysis and the input token.

These measures can be interpreted reasonably directly as embodying competing preferences for *generalization* and *predictive* power, where shorter, simpler grammars tend toward broader constructions that may apply to a wider range of previously unencountered data, while larger, more specific grammars hew more closely to past data and thus permit simpler analyses. The extreme cases serve to illustrate this point: a grammar consisting of a single construction allowing any combination of words will fit all new data, though it may not be very informative about the specific examples likely to be encountered (or their semantic relationships); on the other hand, a grammar consisting of one (maximally specific) construction for each example seen captures the data per-

[1] As noted earlier, both the formalism and the analyzer have been extended to incorporate probabilistic information more fully, as described in Bryant (2008).

fectly but does not generalize well to new situations. The evaluation criteria provide a heuristic for guiding the grammar toward an optimal balance between these extremes.

The remainder of this section defines the grammar and data length measures employed by the model. The approximations used for these sacrifice some information-theoretic precision for the sake of both expository clarity and computational convenience, on the assumption that the model's overall behavior should, given sufficient data, be able to converge on a stable and useful grammar despite these simplifications. That is, any reasonable measure should ultimately lead to improved grammars, with more useful constructions strengthening and less useful ones weakening over time. On a more pragmatic note, the limited amount of data that may be available in general also argues against adopting too precise a measure.

A final caveat: though the evaluation criteria are motivated by the competing usage-based preferences noted above, they are not intended as direct claims about the psychological validity of the particular definitions given. We will revisit this issue in Section 7.5.

7.2.1 Counts and weights

We first review some of the terms used in the grammar and data length criteria. Section 4.1.1 defined *count* and *weight* terms associated with each construction of a grammar:

- The **count** of a construction **c** is its empirical frequency, *i.e.*, the sum over all tokens $d \in D$ of the number of times an instance of **c** appears in d's analysis:

$$\text{count}(\mathbf{c}, D) = \sum_{d \in D} \sum_{x \in \text{analysis}(d)} \delta_{\mathbf{x},\mathbf{c}},$$

where $\delta_{i,j}$ is defined by

$$\delta_{i,j} \equiv \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{c}, \\ 0 & \text{if } \mathbf{x} \neq \mathbf{c}. \end{cases}$$

- The **weight** of each construction **c** normalizes the count over all the constructions in the grammar. This term is useful in the calculation of data length (by virtue of its inclusion in the analysis scoring heuristic defined in Section 4.3):[2]

$$\text{weight}(\mathbf{c}, D) = \frac{\text{count}(\mathbf{c}, D)}{\sum_{\mathbf{x} \in G} \text{count}(\mathbf{x}, D)} \tag{7.2.3}$$

We also track constructional co-occurrence as follows:

[2]It is also useful for approximating a construction's prior probability $\hat{P}(\mathbf{c})$, which plays a role in the detailed (alternative) length criteria discussed in Section 7.4.

- The **constructional co-occurrence count** $[\mathbf{x}, \mathbf{y}]$ is the number of times that instances of constructions $\mathbf{x}$ and $\mathbf{y}$ occur in the same analysis.

- The **constituency count** $[\mathbf{x}.n \leftarrow y]$ is the number of times that an instance of construction $\mathbf{x}$ has its constituent n filled by an instance of construction $\mathbf{y}$. Note that the relative frequency of this constituent-filler pair is then $\frac{[\mathbf{x}.n \leftarrow y]}{[x]}$.

These counts are useful both in calculating data length and in setting initial weights for newly created constructions.

7.2.2 Grammar length

The grammar length $\mathcal{L}(G)$ is a sum over the lengths of all of its constructions:

$$\mathcal{L}(G) = \beta_g \cdot \sum_{\mathbf{c} \in G} \mathcal{L}(\mathbf{c}), \tag{7.2.4}$$

where $\beta_g$is a weight parameter that controls the relative bias toward model simplicity.

The length of a construction depends on the length of specifying both its constituents and parents and the lengths of its form and meaning domain poles (as defined in Section 4.1.1). Treating all constructions in constituent and parent type declarations as having a constant constructional declaration cost ℓ_c:

$$\mathcal{L}(\mathbf{c}) = \ell_c \cdot (|\,\mathrm{constituents}(\mathbf{c})| + |\,\mathrm{parents}(\mathbf{c})|) + \mathcal{L}(\mathbf{c}_f) + \mathcal{L}(\mathbf{c}_m) \tag{7.2.5}$$

The length of a domain pole d similarly incurs a constant cost ℓ_s for each of its (schema) type $\mathrm{declarations}(d)$ and constraints:

$$\mathcal{L}(d) = \ell_s \cdot |\,\mathrm{declarations}(d)| + \sum_{r \in \mathrm{constraints}(d)} \mathcal{L}(r) \tag{7.2.6}$$

The length of a constraint depends on the total path lengths (*i.e.*, slot chain lengths) of its arguments, where a constant cost of ℓ_r is assumed for each constraint and ℓ_p for each path slot:

$$\mathcal{L}(r) = \ell_r + \sum_{a \in \mathrm{args}(r)} \ell_p \cdot |a| \tag{7.2.7}$$

Note that the size of a simple lexical construction with no constituents or parents depends upon just the size of its associated form and meaning domains.

Example. Figure 7.1 illustrates the length metrics just defined using a THROW-TRANSITIVE construction from the familiar object-throwing domain annotated with lengths for its various components. The construction, abbreviated `tt` below, has three constituents and no parent constructions.

Its form domain has two constraints, each of length 2 (1 for each argument), and its meaning domain has two constraints, each of length 3 (where the term t2$_m$.thrower, for example, is a slot chain of length 2). The length of the construction is computed as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{tt}) &= \ell_c \cdot |\,\text{constituents}(\mathbf{tt})| + \mathcal{L}(\mathbf{tt}_f) + \mathcal{L}(\mathbf{tt}_m) \\ &= 3\,\ell_c + 2\,\ell_r + 4\,\ell_p + 2\,\ell_r + 6\,\ell_p \\ &= 3\,\ell_c + 4\,\ell_r + 10\,\ell_p\end{aligned}$$

So, for example, if $\ell_c = \ell_r = \ell_p = 1$, $\mathcal{L}(\mathbf{tt}) = 13$. Generally, ℓ_r and ℓ_p are much smaller than ℓ_c, corresponding to a higher cost of referring to a specific construction than to a particular constraint type or schema role (see Section 7.4).

```
construction THROW-TRANSITIVE ⟹ 3ℓc + 2ℓr + 10ℓp
  constituents
    t1 : REF-EXPR                  ⟹ ℓc
    t2 : THROW-CN                  ⟹ ℓc
    t3 : REF-EXPR                  ⟹ ℓc
  form
    t1f before t2f                 ⟹ ℓr + 2ℓp
    t2f before t3f                 ⟹ ℓr + 2ℓp
  meaning
    t2m.thrower ⟷ t1m              ⟹ ℓr + 3ℓp
    t2m.throwee ⟷ t3m              ⟹ ℓr + 3ℓp
```

Figure 7.1. Example of construction length calculation for THROW-TRANSITIVE

7.2.3 Data length

The data length $\mathcal{L}(D)$ should reflect the total amount of information observed in all input tokens d in D. For current purposes, tokens are assumed to be independent and identically distributed, so that the total length is approximated as the sum of the individual lengths d given the grammar G:[3]

$$\mathcal{L}(D|G) = \beta_d \cdot \sum_{d \in D} \mathcal{L}(d|G) \tag{7.2.8}$$

where β_d= 1 - β_g, *i.e.*, $\beta_d$represents the relative bias toward data encoding simplicity.

[3]That is, the probability $P(d_1|d_x) = P(d_1)$ for all d_x, so their joint probability is the product of each in isolation:

$$P(D|G) = \prod_{d \in D} P(d|G)$$

This of course is a gross oversimplification that ignores the general coherence of conversation. It is perhaps less appalling as a model of the youngest human language learners, who are not ideal Gricean interlocutors and may have a relatively narrow window across which utterances must cohere. This assumption has been relaxed in Mok's (2008a) work addressing contextually grounded constructions (see Section 9.3).

How should the length of an input token be defined? Recall that the grammar is a particular encoding that captures certain redundancies in the data; the length of an input token $d = (u, z)$ can be compressed to the extent that it conforms to these recognized patterns. But the degree of conformity depends not just on the grammar but also on the performance of an analyzer using the grammar. If a grammar happens to have a construction $\mathbf{c_d}$ encoding *precisely* the information of an input token d — that is, all the forms in the utterance u and all the meanings in the context z, including all entities and relations in each domain — then a successful analysis need only identify that (idiomatic) construction to predict all the information in the token. This input token can thus be encoded as an instance of $\mathbf{c_d}$, where its length is the cost of referring to that construction. On the other extreme, if a grammar is relatively impoverished, containing no constructions relevant to the forms and meanings in a given input token, the length of that token is equal to the full cost of encoding all of those forms and meanings in their entirely.

Typically, however, the constructional analysis specifies a set of constructs that together predict a subset of the information in the full utterance-context pair. The remainder represents the degree to which the input token diverges from the expectations encoded by the grammar, *i.e.*, the amount of information in the token beyond that predicted during language understanding. In general, then, the length of an input token depends on the length of its analysis a_d (that is, which constructions a_d instantiates, and in what configuration), as well as the length of encoding any discrepancies between the analyzed form a_f and the utterance u, and between the analyzed meaning a_m and the context z.

Reinvoking the equivalence between optimal code length and inverse log probability, we can recast the length of the analysis $\mathcal{L}(a_d|G)$ in terms of the probability $P(a_d \mid G)$ of an analysis a_d given a grammar G, as defined in the discussion of the *constructional score* in Section 4.3. Exploiting the same approximations used for the constructional score, we approximate the root and constituent probabilities using constructional weights:

$$\mathcal{L}(a_d|G) = -\log P(a_d|G) \tag{7.2.9}$$

$$\approx -\log\left(\prod_{r\in\text{roots}(a_d)} P(r) \prod_{\substack{c\in a_d,\\ c.n\in\text{constituents}(c)}} P(c.n|c)\right) \tag{7.2.10}$$

$$\approx \sum_{r\in\text{roots}(a_d)} -\log P(r) + \sum_{\substack{c\in a_d,\\ c.n\in\text{constituents}(c)}} -\log P(c.n|c) \tag{7.2.11}$$

$$\approx \sum_{r\in\text{roots}(a_d)} -\log(\text{weight}(\mathbf{r})) + \sum_{\substack{c\in a_d,\\ n\in\text{constituents}(\mathbf{c})}} -\log(\text{cweight}(\mathbf{c}.n, \text{type}(c.n))) \tag{7.2.12}$$

where either the uniform or non-uniform version of the $\mathrm{cweight}$ function over constituent filler types can be used. This measure, though an approximation, assesses smaller lengths for high-scoring tokens and larger ones for low-scoring tokens.

To assess lengths for the form and meaning discrepancies, it is convenient to draw upon the results of the contextual resolution process, which determine which parts of the utterance and context are accounted for by a putative analysis. One option is to define a length metric for the unmatched forms and meanings (*i.e.*, the sets referred to as $\bar{A}_f$ and $\bar{A}_m$ in the contextual mapping operations of Section 6.2), similar to those given in Section 7.2.2 for grammar length. A simpler approximation exploits the form and meaning *recall* functions defined in Section 4.3.2; these functions assign higher scores to analyses with relatively fewer such discrepancies, and all scores fall in the range [0,1]. We can thus approximate the overall length of the input token as follows:

$$\mathcal{L}(d|G) = \mathcal{L}(a_d|G) \;-\; \log(\mathrm{recall}_f(d)) \;-\; \log(\mathrm{recall}_m(d)) \tag{7.2.13}$$

The three addends in (7.2.13) effectively measure the information in the analysis itself and the degrees of fit for each of the form and meaning domains, where the latter two can be seen as error-correction terms. As suggested by the parallels between $\mathcal{L}(d|G)$ and the *token score* defined in Section 4.3.3, a smaller length for a token corresponds to a token that is more fully predicted by the current grammar, allowing better analyzer and resolution performance.

Example. Recall the example score calculation from Section 4.3.3, based on analyzing the *throw the ball* input token (Figure 4.4). Assuming the same root weight for the THROW-IMP construction and constituency counts for its constituents (THROW and BALL), the length of the analysis is calculated as follows:

$$\begin{aligned}
\mathcal{L}(a_d|G) &= -\log(\mathrm{weight}(\mathbf{r}_a)) + \sum_{\substack{c\in a_d,\\ n\in\mathrm{constituents}(\mathbf{c})}} -\log(\mathrm{cweight}(\mathbf{c}.n, \mathrm{type}(c.n))) \\
&= -\log(\mathrm{weight}(\mathbf{r}_a)) + \sum_{\substack{c\in a_d,\\ n\in\mathrm{constituents}(\mathbf{c})}} -\log\left(\frac{[\mathbf{c}.n \leftarrow \mathrm{type}(c.n)]}{[\mathbf{c}]}\right) \\
&= -\log(.05) - \log\left(\frac{10}{10}\right) - \log\left(\frac{7}{10}\right) = 4.84 \\
\mathcal{L}(d|G) &= \mathcal{L}(a_d|G) \;-\; \log(\mathrm{recall}_f(d)) \;-\; \log(\mathrm{recall}_m(d)) \\
&= \;4.84 \;-\; \log(.50) \;-\; \log(.432) = 7.05
\end{aligned}$$

7.3 Updating description length

This section specifies how the learning operators defined in Chapter 6 affect the grammar and data length metrics. In general, each operation incurs an increase in the grammar length (equal to the cost of encoding the new construction(s)) and a decrease in the data length (equal to the improved encoding of the example that motivated its application).

7.3.1 Updating grammar length

Calculating the incremental change in grammar length is relatively straightforward, since the grammar length is increased by the length of any newly introduced construction $\mathbf{c}$, plus any change to existing constructions in the grammar due to the introduction of $\mathbf{c}$:

$$\begin{aligned} \mathcal{L}(G') &= \mathcal{L}(G) + \mathcal{L}(\mathbf{c}) + \sum_{\mathbf{x} \in G} \Delta\, \mathcal{L}_{\mathbf{x}}(\mathbf{c}) \\ \Delta\, \mathcal{L}(G) &= \mathcal{L}(\mathbf{c}) + \sum_{\mathbf{x} \in G} \Delta\, \mathcal{L}_{\mathbf{x}}(\mathbf{c}) \end{aligned} \tag{7.3.1}$$

In particular, the merging and splitting operations may decrease the length of the rewritten source constructions (as detailed below).

New constructions must be initialized with an appropriate count. The simplest strategy is to initialize all constructions to 1, on the assumption that useful constructions will be reinforced over time. Alternatively, we can estimate initial counts based on the counts of their source constructions, as discussed below. For simplicity we initialize constructional co-occurrence counts and constituency counts (defined in Section 7.2.1) to 1 (instead of the possibly problematic 0).

Relational mapping The relational mapping operation creates a new construction $\mathbf{c}$ based on shared structure between previously known constructions instantiated in the analysis of an input token; these constructions are constituents of the newly learned construction. The length of the grammar is increased accordingly. Assuming these constituents are of types $\mathbf{a}$ and $\mathbf{b}$, we set the new construction's initial count based on the relevant co-occurrence and constituent-filler counts.

$$\text{count}_0(\mathbf{c}) = [\mathbf{a}, \mathbf{b}] - \left(\sum_{n \in \text{constituents}(\mathbf{a})} [\mathbf{a}.n \leftarrow b] \right) - \left(\sum_{m \in \text{constituents}(\mathbf{b})} [\mathbf{b}.m \leftarrow a] \right) \tag{7.3.2}$$

That is, we assume that the new construction may have been instantiated by previous input tokens involving both constituent constructions, unless one was a constituent of the other.

Merging The merge operation creates a construction $\mathrm{merge}(\mathbf{a}, \mathbf{b})$ based on the shared structure of two existing constructions, and it rewrites $\mathbf{a}$ and $\mathbf{b}$ as subcases of the new mapped construction. The change to the grammar description length depends on the amount of shared structure s between the two constructions:

$$\Delta\,\mathcal{L}(\mathbf{a}) = \Delta\,\mathcal{L}(\mathbf{b}) = \ell_c - s$$

$$\Delta\,\mathcal{L}(G) = s + (\ell_c - s) + (\ell_c - s)$$

$$\Delta\,\mathcal{L}(G) = (2 \cdot \ell_c) - s$$

Each of the rewritten constructions decreases in length by s but incurs the cost of a subcase declaration. Note that this interpretation of the merge operation incurs a net increase in the grammar size, if $s = 1$; as a result, merges tend to be discouraged unless the constructional overlap $s >= 2 \cdot \ell_c$. The count on the new construction is initialized to the sum of the counts of its two subcase constructions.

$$\mathrm{count}_0(\mathrm{merge}(\mathbf{a}, \mathbf{b})) = \mathrm{count}(\mathbf{a}) + \mathrm{count}(\mathbf{b})$$

The counts on the rewritten subcase constructions $\mathbf{a}'$ and $\mathbf{b}'$ remain unchanged.

Joining The join operation exploits competition for a shared constituent between its motivating constructions $\mathbf{a}$ and $\mathbf{b}$. The newly formed construction $\mathrm{join}(\mathbf{a}, \mathbf{b})$ will thus cover cases involving *either* the co-occurrence of $\mathbf{a}$ and the unshared constituent of $\mathbf{b}$ (of type $\mathbf{b_u}$), *or* the co-occurrence of $\mathbf{b}$ and the unshared constituent of $\mathbf{a}$ (of type $\mathbf{a_u}$). We take the average of the relevant co-occurrence counts:

$$\mathrm{count}_0(\mathrm{join}(\mathbf{a}, \mathbf{b})) = \frac{[\mathbf{a}, \mathbf{b_u}] + [\mathbf{b}, \mathbf{a_u}]}{2}$$

Splitting The split operation produces a new construction $\mathrm{split}(\mathbf{a}, \mathbf{b})$ that constitutes the difference between the (larger) construction $\mathbf{a}$ and its subpart $\mathbf{b}$; $\mathbf{a}$ also gets rewritten to include the new construction as a constituent. The slight change in grammar length is independent of the amount of shared structure s:

$$\Delta\,\mathcal{L}(\mathbf{a}) = \ell_c - s$$

$$\Delta\,\mathcal{L}(G) = s + (\ell_c - s)$$

$$\Delta\,\mathcal{L}(G) = \ell_c$$

That is, the information in s is simply moved from $\mathbf{a}$ to the new construction. The constant length increase is thus due entirely to the added constituent declaration on $\mathbf{a}$. The count of

$\text{split}(\mathbf{a}, \mathbf{b})$ is set to be same as that of $\mathbf{a}$, on the assumption that it would have been identified in all of those occurrences.

$$\text{count}_0(\text{split}(\mathbf{a}, \mathbf{b})) = \text{count}(\mathbf{a})$$

The count of the original larger construction $\mathbf{a}$ remains unchanged.

7.3.2 Updating data length

Calculating the change to data length poses a more significant challenge, since new constructions might affect the analyses of many input tokens in the training corpus. The standard Bayesian strategy of reanalyzing every training token is infeasible for both computational and cognitive reasons: it is inefficient to do this for any but the smallest corpus sizes, and it is cognitively implausible that specific input utterances and contexts are effectively and indefinitely memorized.[4] Indeed, untrammeled access to one's entire history of linguistic experience would tend to undermine the learner's motivation to compress the data.

Alternatively, the model can retain a subset of encountered tokens as a more reasonably sized test corpus whose change in data length is taken to be representative of that of the entire corpus. This subset might be considered the model's *internal* development set against which to assess candidate operations. These might be chosen based on various computationally and cognitively motivated criteria:

- *Recent* tokens, *i.e.*, those occurring within a given history window, may be the most representative of the current (and imminent) input; they can be considered primed or otherwise still in short-term memory.
- *Challenge* tokens whose analyses are judged as poorly understood by the current grammar are retained as salient candidates for improvement. These include tokens with scores below a threshold level, partial analyses (having no single root construct) or ambiguous analyses (having multiple analyses scoring within a maximum window of the best analysis).
- A *random* subset of tokens may be the most representative of the full range of data encountered. Cognitively this might be taken to encompass the many other factors not explicitly included in the model that might make some experiences particularly salient relative to others, such as the achievement (or failure to achieve) an especially worthy goal or some other emotionally charged circumstance.

[4] Or, as expressed by Steven Wright: "You can't have everything. Where would you put it?"

Costly reanalysis can be further limited by noting that many operations may affect a small number of the tokens within the internal test set, *e.g.*, only the ones containing the relevant lexical items or construction types.

The experiments described in the next chapter combine the strategies above: subsets consisting of recent tokens, challenge tokens and randomly chosen tokens are each recorded with their best analyses. Candidate learning operations trigger the re-evaluation of targeted subsets of these corpora affected by the newly created constructions.

7.4 Incorporating reference costs

The length heuristics defined so far make a number of simplifications for computational and expositional ease. This section considers a refinement of the evaluation criteria that incorporates a more precise view of the cost of referring to a structure. Recall that a complex construction may assert various kinds of constraints that make reference to other constructions. Likewise, an analysis involves constructs that refer to the constructions they instantiate. The length measures defined so far have assumed that all references are equal cost.

In view of the connection between probability and code length noted earlier, however, the length or cost of a *reference* to a particular construction or schema could depend on its frequency, such that structures that are more frequent or familiar are less costly to reference. Similarly, different constraint types (role restriction, identification, filler, evokes) could be defined to have different costs, also proportional to empirical frequency, assigning a lower cost to frequently occurring constraints (*e.g.*, word order, or role identification) than to rarely occurring ones (*e.g.*, vowel shifting, inversion). As discussed in Section 7.1, optimal code length is equivalent to the inverse log of the construction's prior probability. This probability can be estimated based on empirical frequency — in other words, it is conveniently defined in terms of the weight function from (7.2.3). We define the **pointer** (or **reference**) cost of referencing a construction $\mathbf{c}$ as the inverse log of $P(\mathbf{c}, D)$:

$$\begin{aligned}\mathrm{cost}(\mathbf{c}) &\approx -\log P(\mathbf{c}, D)\\ &\approx -\log \mathrm{weight}(\mathbf{c})\\ &\approx -\log \frac{\mathrm{count}(\mathbf{c}, D)}{\sum_{\mathbf{x}\in G}\mathrm{count}(\mathbf{x}, D)}\end{aligned} \tag{7.4.1}$$

Incorporating construction reference costs requires the following changes to the length metrics:

- The length of a *construction* $\mathbf{c}$ includes reference costs for any constituent and parent constructions, replacing the constant cost (reflecting uniform constructional probability) ℓ_c used in (7.2.5):

$$\mathcal{L}'(\mathbf{c}) = \sum_{n \in \text{constituents}(\mathbf{c})} \text{cost}(\text{type}(n)) + \sum_{p \in \text{parents}(c)} \text{cost}(p) + \mathcal{L}'(\mathbf{c}_f) + \mathcal{L}'(\mathbf{c}_m)$$

- The length of a *domain pole* d (*i.e.*, the form pole $\mathbf{c}_f$ or meaning pole $\mathbf{c}_m$) includes reference costs for its type declarations, based on a similar change to (7.2.6):

  $$\mathcal{L}'(d) = \sum_{t \in \text{declarations}(d)} \text{cost}(\text{type}(t)) + \sum_{r \in \text{constraints}(d)} \mathcal{L}'(r)$$

  where $\text{type}(t)$ of a type constraint or evokes statement t is the construction or schema type to which it refers.

- The length of a *constraint* r depends on the frequency of the specific constraint type $\text{type}(r)$, rather than the constant cost ℓ_r used in (7.2.7):

  $$\mathcal{L}'(r) = \text{cost}(\text{type}(r)) + \sum_{a \in \text{args}(r)} \ell_p \cdot |a|$$

Updating with reference costs

Changes to the grammar based on the addition of new construction must also take changes to reference costs into account: Each operation adds a new construction with initial counts, which increase the total grammar counts and therefore decrease each existing construction's weight, in accord with the normalization of (7.4.1). The change to the grammar length $\Delta\,\mathcal{L}(G)$ from (7.3.1) for a new construction $\mathbf{c}$ must therefore also be modified to reflect changed reference costs, where $\text{cost}(G)$ is the (reference) cost of the entire grammar:

$$\Delta\,\mathcal{L}'(G) = \mathcal{L}(\mathbf{c}) + \Delta\,\text{cost}(G) \tag{7.4.2}$$

$$\Delta\,\text{cost}(G) = \sum_{\mathbf{x} \in G} (\Delta\,\text{cost}(\mathbf{x}) \cdot |\,\text{refs}(\mathbf{x})|) \tag{7.4.3}$$

$$\Delta\,\text{cost}(\mathbf{x}) = \text{cost}_{G'}(\mathbf{x}) - \text{cost}_G(\mathbf{x}) \tag{7.4.4}$$

$$= -\log \text{weight}_{G'}(\mathbf{x}) - \text{cost}_G(\mathbf{x}) \tag{7.4.5}$$

$$= -\log \frac{\text{count}(\mathbf{x})}{\text{count}(G')} - \text{cost}_G(\mathbf{x}) \tag{7.4.6}$$

$$\text{count}(G') = \text{count}(G) + \text{count}_0(\mathbf{c}) \tag{7.4.7}$$

That is, assuming each construction $\mathbf{x}$ is referenced throughout the grammar G a total of $|\,\text{refs}(\mathbf{x})|$ times, the change in grammar cost $\Delta\,\text{cost}\,G$ is calculated in (7.4.3) by summing over the total changes in the cost to each construction. These individual changes in cost between the original and updated grammar (referred to in (7.4.4) as $\text{cost}_G(\mathbf{x})$ and $\text{cost}_{G'}(\mathbf{x})$, respectively) can be calculated based on $\text{weight}_{G'}(\mathbf{x})$ (its weight in the updated grammar G') and the relevant new and original counts, where $\text{count}(G')$ in the denominator of (7.4.6) is the total count over all constructions in the updated grammar G' (*i.e.*, the original count incremented by the initial count $\text{count}_0(\mathbf{c})$ of the new construction $\mathbf{c}$).

7.5 Discussion

The barrage of definitions and evaluation criteria offered in this chapter should not obscure their basic motivation: to provide the learning model with a means of deciding what moves to make through the space of possible grammars, and to do so on the basis of *simplicity*. The purpose of each of the length definitions given is to encourage the most compact encoding of both the grammar and the data. They are thus in some sense intended as objective measures of information content. But it should be clear that they are nonetheless at least indirectly tied to domain-specific factors: the nature of the hypothesis space, as defined by the ECG formalism, motivates and constrains the definition of grammar length, and the goals of language understanding, as defined in our simple models of analysis and resolution, are directly transferred to the definition of data length.

Several other language acquisition models in the computational literature have employed criteria based on MDL. Goldsmith (2002) describes a model for learning concatenative morphological segmentation with measures similar to those used here, though concentrating on the form domain (and restricted to orthographic strings), and Wolff (1982) describes an early model of language acquisition whose learning operations are designed to remove redundancy in the grammar. While MDL was not used explicitly as the determining factor governing these uses, the operations of the model indirectly led to the same tradeoff between size and expressive power captured here. More recently, Chater (2004) has argued for a simplicity principle as the basis for addressing issues of learnability, and shown that under certain conditions an MDL-based framework provides a useful basis for avoiding overgeneralizations (Onnis *et al.* 2002) despite receiving only positive examples. (See Section 9.2.1 for additional discussion of related approaches.)

Happily, a simplicity-oriented learner also fits well with proposals in the developmental literature. Clark (2003) identifies several principles for word learning that are directly realized by aspects of the evaluation criteria defined here:

- *Simplicity of form* corresponds to the length of a construction's form domain $\mathcal{L}(\mathbf{c}_f)$; the criteria used here extend simplicity to include the domain of meaning as well.

- *Transparency of meaning* corresponds in this framework to minimizing constraints that must be expressed independent of constituent constructions. That is, meanings of entire units based on the meanings of known subparts are preferred.

Slobin's (1985) operating principles also include general injunctions to pay attention to frequencies of usage and co-occurrence, which correspond directly to the inclusion of weights and co-occurrence counts in the model.

The resonance of the current model with these proposals suggests that simplicity as a basis of learning is promising on both domain-general and domain-specific grounds. As with the learning operations, the model is designed to allow a potentially large (if not infinite) set of variations upon the themes explored by our defined criteria; I have in this chapter but skimmed the surface, making many simplifying assumptions along the way. But the power of the model should — in theory — derive from the fact that neither the set of operations nor the evaluation criteria need be exhaustive, definitive or even infallible. They need only be reasonable enough to guide the learner toward more useful parts of the hypothesis space. The next chapter aims to demonstrate that, all things being reasonable, reasonable results ensue.

Chapter 8

The model in action

However beautiful the strategy, you should occasionally look at the results.
— Winston Churchill

This chapter puts the model to work: I present results of applying the model to learn early English relational constructions in the motion event domain. Section 8.1 summarizes the methodology applied and the experimental conditions for our simulated learner. Two learning experiments are then described, both qualitatively and quantitatively, in Sections 8.2 and 8.3.

8.1 Nomi's world

Despite the ubiquity of appropriate test subjects in our stage of interest, the most feasible (ethical) means of testing the proposed model is by way of a computer simulation of the learner's situation. While such simulations cannot encompass the full extent of the human learner's experience, an artificially constructed environment has a significant advantage in transparency: all assumptions with respect to the learner's prior knowledge, the inputs to the system, and the internal mechanisms driving language learning and use must be made explicit. Computational modeling also forces the experimenter to selectively abstract away from details that may be inessential for the class of problems of interest; the forced suppression of detail is a crucial scientific tool for coping with the extraordinary complexity of human cognition.

As should be clear from the subject matter, as well as the many simplifying assumptions proffered already, the current task poses significant challenges to computational modeling, both theoretical and practical. These challenges stem in part from the maximally inclusive nature of the approach taken here: the model is intended to conform with evidence from all disciplinary quarters, including especially the inferential and pragmatic prowess attributed to young children in our stage of interest. Such skills ease some aspects of the language learning task, but they present a modeling challenge in their own right, even without trying to integrate them into a theory of language learning.

More generally, the emphasis on meaning and context in the current model falls prey to the definitional concerns of Bloomfield's noted in Chapter 2: while ample corpora of speech- and text-based data are available, very little of it includes the kinds of semantic information presumed in the current model to be the driving force in language acquisition. As noted earlier, the input data to the model is based on the CHILDES database, a repository of parent-child interactions (MacWhinney 1991) spanning a wide range of languages, ages and media and providing a rich source of primary data. Most of it is not, however, particularly well-suited for the purposes of computational modeling. One major challenge stems from the fact that many child language transcripts focus on the productive capacity of the child, rather than on utterances (*i.e.*, the input data) from the parent, as is more relevant for the current model's focus on comprehension as the route to learning. Another challenge is finding appropriate data from the relevant stage of learning addressed by the model, *i.e.*, the transition from single-word utterances into multi-unit relational constructions. Most significantly, however, transcript data typically provides only syntactic, morphological and phonological information; semantically and contextually available information is only haphazardly mentioned.

Based on these restrictions, the experiments described here are intended mainly to demonstrate concretely how the general problem assumptions given in Chapter 5 and the specific technical constraints given in Chapters 6 and 7 can be put into practice, emphasizing qualitative behaviors and the potential of the learning model to scale to more realistic settings. The remainder of this section reviews the methodological procedures and overall experimental framework used in the experiments described in subsequent sections.

8.1.1 Input assumptions

The input tokens specified for learning, defined in Chapter 4, include much more semantic and pragmatic information than is typically available in the CHILDES database. Longitudinal corpora

Age	Parent	Child
1;2.29	whoopsi daisy	
	it fell right off didn't it?	
	it fell right off.	
	don't you put that on my hair you.	
	no # don't you put that on my hair.	eh.

Table 8.1. Example sequence from the Naomi corpus (Sachs 1983); the situationally inferrable referent of *it* and *that* is a popsicle mentioned earlier in the dialogue.

do, however, provide a useful indication of the range of utterances encountered by a child at the appropriate age. The experiments undertaken here use input based on the Sachs corpus (Sachs 1983). The transcript data consists of parent and child utterances occurring in the context of a joint background activity (*e.g.*, a meal or play situation) involving the English-speaking child Naomi (referred to in the corpus mostly as *Nomi*), spanning approximately age 15 to 24 months. During the early sessions, the parents do most of the talking, interspersed occasionally by single-word contributions from Naomi; by the later sessions she has grown into a seasoned discourse participant. The Sachs corpus is representative of other corpora in that it covers the standard (fairly limited) repertoire of child-oriented playtime scenes. Some example exchanges were given in Chapter 1 (Table 1.1); Table 8.1 shows another representative sequence of utterances from early in the corpus (1;2.29).

As this sequence illustrates, child-directed utterances are rife with situationally dependent expressions (*e.g.*, what *it* is, what is being put on the speaker's hair), which are not always preceded by explicit discourse referents. They differ markedly from standard corpora used in computational linguistics applications, and they would pose a considerable natural language challenge even if learning were not part of the task. But while parent-child exchanges may lack the coherence of adult conversation, they do exhibit *situational* continuity through the general flow of events — in this case, the fall and subsequent (objectionable) placement of a popsicle mentioned earlier in the dialogue. As a result, most situational referents can be easily inferred even in the absence of explicit discourse referents.

A subset of the corpus was annotated with basic scene information for motion-related events. Scene annotations include the speaker's inferred communicative intent, as well as available referents in the immediate discourse and situational context, approximating the parent-child common ground in the sense of Clark (1992). The annotations include information about the basic scene type (related to the prototypical scenes of Slobin (1985), mentioned in Chapter 2) and the main scene participants, typically including the entity engaging in motion, the path or goal of the motion and (if

Spk	Utterance	Annotation
MOT	it fell right off.	M fall/off COM=3 VstP: m1:motion[mover=popsicle goal=off]
MOT	# do-'nt you put that on my hair you.	M put/on my hair NIMP=2 VcPG cm1:cm[(causer=2) mover=popsicle goal=hair]
MOT	you fell down?	M fall/down Q/CONF-2 VsP: m1:motion[mover=2 path=down]
MOT	# whoops Goldie fell out.	M fall/out COM=3 VstP: m1:motion[mover=goldie goal=out]
MOT	do-'nt throw the bear.	M throw NIMP=2 Vc cm1:cm[(SIT causer=2) mover=bear]

Table 8.2. Examples of annotated utterances extracted from input corpus, showing the speaker, utterance and the token annotation. The scene annotation in the first line includes the main and secondary predicates in the utterance, speech act type and other general scene information. The extended semantic annotation in the second line specifies scene role-fillers, where parenthesized fillers are inferred from the discourse or situational ('SIT') context.

applicable) the participant causing the motion. Sample utterances from the corpus are shown in Table 8.2, along with their scene annotations.[1]

Figure 8.1 shows the input token generated based on the first annotation example in Table 8.2, in the feature structure format introduced in Chapter 4. The annotation codes (shown in the right column of Table 8.2) specify a variety of scene features. In brief, the first line of the annotation specifies the speaker (the mother), a main and secondary predicate (*i.e.*, relational, functional or predicating lexeme), in this case *fall* and *off*, respectively; the speech act (here, 'COM' indicates a comment; other speech act types include 'IMP' imperative, 'NIMP' negative imperative, 'Q' question, and 'CONF' confirmation); the temporality (marked '=' to indicate reference to the ongoing scenario; '-' and '+' indicate past and future events, respectively); the person or perspective taken on the event (where the '3' indicates a third-person perspective); and the motion scene involved (here, a self-motion of an inanimate object with respect to a deictic direction).[2] The second line specifies the role bindings for the relevant motion schema. Here, the mover is specified as a (contextually available) 'popsicle', and the goal of motion is 'off'. (The filler '2' seen in other examples refers to the current addressee.) Together these annotations allow the translation of the raw data above into the input token format suitable for the learning model.

[1]The initial scene annotations (corresponding to the first line of each annotation in Table 8.2) come from a study of crosslinguistic motion expressions in child-adult discourse conducted by Slobin and his students (May *et al.* 1996). This coding schema was supplemented with an extended annotation scheme for specifying semantic role relations, situational reference and episodic information more precisely (Mok 2008b) and applied to the Sachs corpus subset by Josef Ruppenhofer, Eva Mok and the author.

[2]Other motion event types covered by the annotation format include caused motion of an animate or inanimate object, motion caused by an inanimate object, dressing movements involving clothing, and transfer events (May *et al.* 1996).

$$
\left(
\begin{bmatrix}
\text{Utterance} & \\
\text{text :} & \text{“it fell right off”} \\
\text{intonation :} & \text{falling}
\end{bmatrix},
\begin{bmatrix}
\text{Context} & \\
\text{participants :} & \text{Mother, Naomi, pZ:Popsicle} \\
\text{scene :} & \begin{bmatrix}\text{Fall} & \\ \text{mover :} & \text{pZ} \\ \text{goal :} & \text{tlZ}\end{bmatrix}, \begin{bmatrix}\text{tlZ:Trajector-Landmark} & \\ \text{trajector :} & \text{pZ} \\ \text{landmark :} & \\ \text{relation :} & \text{off}\end{bmatrix} \\
\text{discourse :} & \begin{bmatrix}\text{DiscourseSpace} & \\ \text{speaker :} & \text{Mother} \\ \text{addressee :} & \text{Naomi} \\ \text{speech act :} & \text{comment} \\ \text{joint attention :} & \text{pZ} \\ \text{activity :} & \text{play} \\ \text{temporality :} & \text{ongoing}\end{bmatrix}
\end{bmatrix}
\right)
$$

Figure 8.1. Sample input token based on the utterance “it fell right off”

One important class of information that is to some degree lost in this translation is the aforementioned connection to the ongoing flow of discourse. That is, the annotation scheme just described includes a relatively complete snapshot of the information needed for analysis and learning to proceed, which allows each token to be treated independently of the rest. This simplifying assumption is motivated in part by the fact that the labeled motion utterances are not generally contiguous in the corpus; it is also assumed that children at early stages of development generally do not exploit an extended contextual history.[3]

Motion predicates				**Particles**	
put	184	push	13	on	104
get	106	pick	12	in	75
go	97	spill	11	down	60
take	69	run	11	out	50
come	57	move	9	up	46
fall	36	drop	9	off	41
throw	24	climb	9	back	33
bring	20	slide	8	over	26
pull	19	pour	8	to	22
jump	16	swim	7	away	16

Table 8.3. Highest frequency motion predicates and particles from annotated subset of the Sachs corpus, shown here with their counts in the annotated adult utterances.

Table 8.3 lists the most frequent main and secondary predicates occurring in the annotated Naomi input data. In general these correspond, respectively, to motion predicates (both relatively

[3]Mok (2008a) describes a language learning model that relaxes this assumption to include an explicit model of discourse and situational history, as also described in Chang & Mok (1998).

general, such as *put* and *go*, and more concrete, such as *fall* and *throw*) and particles indicating locative relations and resultative states. These categories provide an illustrative sample of the semantic range covered by the data.

8.1.2 Initial lexicon and schema set

The initial schema set contains basic entities, actions and locations familiar to the child by the time of the two-word stage, as specified in Chapter 5. Conceptual frames, such as possession, action, motion and location are organized in a multiple inheritance hierarchy as defined in Section 4.1.1.

The initial lexicon contains labels for the people, things, locations and actions familiar to the child. A general constructional category of LABEL is a default pairing of a word form with some meaning. The experiments to be described also included subdivisions of NLABEL, VLABEL and TLLABEL. These categories are mnemonic for noun-like, verb-like and trajector-landmark classes, and they are motivated as (labels for) the conceptual categories of entities, actions and spatial relations. The model does not assume these are fixed or inherent parts of speech; rather, these rough categories are taken to have arisen prior to the current learning scenario, through lexical learning based on conceptual similarity.

8.1.3 Learning model parameters

The learning model as described allows for a variety of options in both the search for new operations and the evaluation criteria applied. Assumptions made for the current experiments are as described in Chapter 6 and Chapter 7, along with the specifications below.

Analysis and resolution

- The maximum utterance length allowed is 15.
- The calculation of form and meaning coverage applies a weight ratio of 2:1 between units and relations, on the assumption that, for example, unmapped words may be more salient, and costly, to omit from an analysis than unmapped relations. The same holds for meaning schemas: an unmapped entity or action may be more salient than a role-filler binding.

Operations

- *Relational mapping:* Relations that capture pseudo-isomorphic structure across two already known constructions (in the sense defined in Chapter 6) are preferred. These are further constrained to be *root* constructs within the current analysis. Relations involving salient items in the discourse space are also allowed. The maximum path length $\mathrm{distance}_p$ for constraints (defined in Section 6.2.2) is 3.

- *Merging:* Merging operations are relatively conservative, with $\mathrm{distance}_a = \mathrm{distance}_{max} = 2$. That is, constructional and constituent merges are permitted only for constructional (or meaning) types that are in a parent-child relationship or are siblings (*i.e.*, share a parent).

- *Joining:* Joining is allowed when there are two constructions in the chart with overlapping material of at most one shared construction; the overlapping material must be a single constituent of the same type for each of the two joining constructions.

- *Splitting:* Input tokens with unknown stressed forms license a mapping between a single unknown form and a single unmapped schema.

Evaluation

- The grammar length metric uses uniform costs for constructional references, and the data length metric uses uniform probability constituency counts.

- The data length evaluation is calculated over a subset of the data, as discussed in Section 7.3.2, with a maximum of 10 tokens drawn from the recent, challenge and random subcorpora.

- The calculation of total description lengths applies a weight ratio of 1:2 between the grammar length and the data length.

8.1.4 Basic methodology

The experimental methodology reflects both the problem definition and proposed solution set out in Chapter 5. The learning system is equipped with the initial lexical grammar described in Section 8.1.2. Input tokens are split into *training* (80%) and *testing* (20%) subsets. The basic training procedure follows that outlined in Figure 5.3 from Chapter 5, with training interrupted periodically to evaluate the current learned grammar's performance on the test set. The test set is also referred to below as the *external* test set, since it is used for evaluation external to the learning algorithm. The *internal* evaluation set refers to the subset of the training set available at any given time (internally) to the learner, *i.e.*, the internal dataset used for evaluating and choosing among candidate operations.

Quantitative evaluation metrics primarily track the improvement of the model's comprehension as training progresses. In addition to the performance criteria described in Section 5.2.5, the length criteria defined in Chapter 7 (for the grammar length, data length and total description length) measure the overall compactness of a learned grammar and its associated encoding of the data (either the external test set or the internal evaluation set). Qualitative considerations focus on what kinds of constructions are learned, how these affect comprehension and whether they correspond to the kinds of constructions learned by children at a comparable stage.

8.2 A toy story

The initial experiment is based on a subcorpus FT of the Nomi data similar to the illustrative examples seen so far, consisting of 40 tokens based on the verbs *fall* and *throw*. This restricted set provides a somewhat less chaotic view into the dynamics of the learning system than larger corpora and illustrates several operations at work.

```
ft.1-1: fell down again : Fell, Down, AgainU [61.4 (F 72.7,M 50.0)]
  OP 1: MAP (Fell, Down)    ===>    Fell-Down[dG=38.0 dD=-15.1 dGD=22.9]

ft.1-2: you throw a ball : You, Throw, AU, Ball [52.3 (F 54.5,M 50.0)]
  OP 2: MAP (Throw, Ball) ===>    Throw-Ball[dG=38.0 dD=-15.2 dGD=22.8]

ft.1-3: dont throw the bear : DontU, Throw, TheU, Bear [50.3 (F 54.5,M 46.2)]
  OP 3: MAP (Throw, Bear)    ===>    Throw-Bear[dG=38.0 dD=-15.2 dGD=22.8]
```

Figure 8.2. Trace of learner operating on first few tokens of the FT corpus

Figure 8.2 shows a compressed trace of the program running on the first few tokens. The output shows the incoming tokens, a summary of the results of analyzing them, and a summary of the operation performed, if any. Here, all three tokens have only partial analyses, based on known and unknown lexical items (*e.g.*, from token ft.1-1 above, `Down` and `AgainU`, where the appended `U` indicates an unknown form). The weighted token score is shown, along with the respective form and meaning coverage scores. In all three cases, learning yields a contextual mapping operation, shown with the types of its constituent constructions. The first example produces a `Fell-Down` construction; the bracketed numbers (in the operation line) indicate the (prospective) changes to the grammar, data and total description lengths, respectively. Each operation incurs an increase in the grammar length (due to the addition of the proposed construction), a decrease in the data length (due to the reduced cost of encoding the internal evaluation set when the grammar includes the proposed construction), and a slight increase in the total description length.

Omitted in this compressed display is an indication of the meaningful context and content in both the input and output. The first token's output is expanded in Figure 8.3 to include these. The token's annotation, for example, indicates a downward motion event whose mover is `goldie`, a known (and named) stuffed dog toy in the learner's ken — although not specified by the utterance, the annotation notes it as a situationally available role-filler. The details listed with the analysis indicate the meaning types that are part of the semspec and context, as well as the best-scoring (total) resolution map, with two resolution bindings. Finally, the mapping operation proposed is shown with an abbreviated notation for the two mapped constraints: an ordering relation `fe meets d` and an identity binding `fe.m.goal`, where `fe` and `d` are the local constituent names for the `Fell` and `Down` constructions, respectively.

```
ft.1-1: fell down again  <=====> |M fall/down again COM=3 VstP:A: (toy)
                                  m4:motion[(SIT mover=goldie) path=down]
   ANALYSIS: [Fell, Down, Again], score=[61.4 (F 72.7,M 50.0)]
   -semspec: SS0: d1:Down, fa1:Fall
   -context: fa2Z:Fall, DS1Z:DiscourseSpace, n1Z:Naomi, d2Z:Down, mo1Z:Mother, t7Z:Toy
   -resmap (total, 2/2, score=0.525): SS0.fa1 <=> mo1Z:Fall, dC1.m <=> d2Z:Down
   PROPOSED 1 operation(s):
      1: MAP (Fell, Down)      ===>     Fell-Down[dG=38.0 dD=-15.1 dGD=22.9]
            MAP (d: Down, fe: Fell): fe meets d  ===== d.m <--> fe.m.goal
  OP 1: MAP (Fell, Down)   ===>     Fell-Down[dG=38.0 dD=-15.1 dGD=22.9]
```

Figure 8.3. Detail of learner operating on token ft.1-1

This level of detail suffices to illuminate the mapping process at work in the first token; the next two are similarly straightforward examples of mapping a word order relation to a scene binding, both for a throwing scene. (Note that the unknown words *a* and *the* are bypassed in both cases to allow constructional mappings over two recognized words.) In fact, mapping operations dominate the early training. At token ft.1-10, however, a merge operation is proposed, as shown in Figure 8.4. In this case, a just-learned `Fall-Down` construction is constructionally aligned with the previously learned Fell-Down construction. That relational merging prompts the subsidiary constituent merging of their respective falling-based constructions into a `FallL` constituent category.[4]

```
ft.1-9: fall down : Fall, Down [67.9 (F 85.7,M 50.0)]
  OP 13: MAP (Down, Fall)  ===>  Fall-Down[dG=38.0 dD=-15.2 dGD=22.8]

ft.1-10: fall down : Fall-Down [ Fall, Down ] [79.2 (F100.0,M 58.3)]
  OP 14: MERGE (Fall-Down, Fell-Down)  ===> FallL-Down[dG=59.0 dD=0.0 dGD=59.0]
```

Figure 8.4. Merge of `Fall-Down` and `Fell-Down`, resulting in the `FallL` constituent category

As training continues, the learner hypothesizes further maps in response to the input, including a `Kangaroo-FallL-Down` construction that has an internal `FallL-Down` construction —

[4]Here the `L` is mnemonic for the lifted nature of the new category. In this case, Fell and Fall have meaning poles of the same type, a Fall action, causing an inductive leap generalization to include any construction with that meaning pole.

exhibiting, perhaps, nascent subject-predicate structure. The second iteration through the data yields further merges, such as the ones shown in Figure 8.5.

```
ft.2-8: they 're throwing a ball: They, ReU, Throwing-Ball [ Throwing, Ball ],
                             {skipped: ReU, AU}  [59.7 (F 61.1,M 58.3)]
  PROPOSED 2 operation(s):
   * 10: MERGE (Throwing-Frisbee, Throwing-Ball) ===>
                              Throwing-BallFrisbeeE[dG=66.0 dD=0.0 dGD=66.0]
       6: MERGE (Throwing-Ball, Throw-Bear)      ===>
                              ThrowL-BallBearE[dG=89.0 dD=-12.2 dGD=76.8]
[...]

ft.2-13: throwing the frisbee : Throwing-Frisbee [ Throwing, Frisbee ]
                              {skipped: TheU}  68.1 (F 77.8,M 58.3)
  OP 9: MERGE  (Throwing-Frisbee, Throw-Bear) ===>
                              ThrowL-BearFrisbeeE[dG=89.0 dD=0.0 dGD=89.0]
[...]

ft.2-27: throwing a ball : Throwing-BallFrisbeeE [ Throwing, Ball ]
                              {skipped: AU}  68.1 (F 77.8,M 58.3)
  OP 52: MERGE  (ThrowL-BearFrisbeeE, Throwing-BallFrisbeeE) ===>
                              ThrowL-ToyL[dG=66.0 dD=-19.2 dGD=46.8]
```

Figure 8.5. Later merge operations based on various toy-throwing configurations

The merges proposed and effected in these examples show several different constituent categories in early stages of formation. At token ft.2-8, two merges are proposed, each merging different constituent pairs: operation 10, which would produce a `BallFrisbeeE` constituent category (where the `E` is mnemonic for enumerated) with just two subcases, `Ball` and `Frisbee`; and operation 6, which would produce a similar enumerated `BallBearE` category (along with a lifted category `ThrowL`). The former operation incurs a slightly smaller increase in cost and is chosen. Ultimately, however, the various toy-based categories pattern together, as evidenced by the later operations. First, the merge at ft.2-13 produces the enumerated category `BearFrisbeeE` (along with the lifted category `ThrowL`). Then, at ft.2-27, the proposed merge aligns constituents whose types are, respectively, `BearFrisbeeE` and `BallFrisbeeE`; the merging of these constituent categories warrants an inductive leap to include any construction whose meaning is a subcase of Toy. The resulting `ThrowL-ToyL` construction exhibits generalization beyond seen data, approaching that observed in verb island constructions.

8.2.1 Qualitative observations

As illustrated by the learning sequence above, the model is able to acquire a variety of relational constructions based on the input encountered. While the toy corpus used for this experiment may not shed much light on the long-term dynamics of the system, a few key behaviors suggest the potential for the emergence of more complex structures characterizing grammatical knowledge.

Constituent structure. The model achieves the basic goal of inducing relational, constituent structure from input that is (in surface form, at least) relatively flat and unstructured. These structured mappings represent a key conceptual and representational advance over the simple mappings that characterize much of lexical learning.

Limited generalization. In its initial stages the model mainly applies the relational mapping operation to learn item-specific, concrete word combinations like those earliest acquired by children. As learning progresses, merging operations first lead to minimally variable expressions (similar to those in pivot grammars) with a few alternative event participants, which over time evolve into increasingly inclusive constituent categories. The simple experiment demonstrated above allowed relatively conservative merging; an adjustment of the parameter controlling maximum merging distance would allow more permissive merging over a larger part of concept space. (An increase in this parameter over the course of learning could reflect the development of increasingly general event structure schemas.) Such constructions would approximate verb island constructions and other constructions that are partially item-specific and partially general.

Category formation. Both limited-scope "enumerated" constituent categories (like `BallFrisbeeE`) and (potentially) wider-ranging "lifted" constituent categories (like `FallL` and `ToyL`) are acquired by the system as a result of generalization. In this case, the two kinds of categories do not differ much. In theory, however, the power to make inductive leaps that extend the reach of a given construction is characteristic of, and necessary for, more general argument structure constructions.

Structural ambiguity. A natural consequence of the emergence of structure is the birth of ambiguity. The model works in a more or less bottom-up fashion to posit new relational structures; over time, similar surface forms lead to constructions with different internal constituency structure. For example, at different stages during the learning trial, both `You-Throw` and `Throw-Ball` constructions are proposed, allowing multiple analyses of the sentence *you throw the ball*. Later in learning (though not shown above), a join operation proposes a `You-Throw-Ball` construction with a flat structure over three sibling constituents.

Distributional information. Figure 8.4 shows the number of constructions learned based on each class of predicates.[5] As noted above, these are dominated for this corpus by concrete, lexically specific constructions, with a few more general constructions acquired as learning progresses.

[5] The "other" category includes some constructions learned based on the secondary, directional predicates, such as the `In-Basket` construction and `Books-Down` construction. These constructions potentially presage more general prepositional phrase and particle constructions, respectively.

The two verbs produce similar numbers of lexical constructions, but *throw* has four generalized constructions, compared to two for *fall* — this despite the greater frequency of *fall* tokens (24) over *throw* tokens (16) in the corpus. Examining the data, however, it is clear why this happens: *fall* nearly always appears with either *down* or *over* (in the toppling sense), while *throw* is more liberally applied across potential projectiles. The importance of distributional information in constructional dynamics and the emergence of generalized structures has long been observed (Maratsos & Chalkley 1980; Lieven *et al.* 1997); these phenomena are both consistent with and easily explained by the mechanisms proposed by the current model.

	concrete lexical	variable	total
throw	9	4	13
fall	8	2	10
other	3		3
total	20	6	26

Table 8.4. Number of concrete and variable constructions learned from FT corpus

8.2.2 Quantitative trends

The various metrics defined in the preceding chapters provide a means of tracking the model's progress. The tables in Figure 8.6 summarize the model's performance on the test set before and after learning on two kinds of metrics: the comprehension criteria defined in Chapter 4 and the length criteria defined in Chapter 7. These values are shown along with "gold" standard values, taken from a hand-crafted grammar designed to cover most of the simple nominal, verbal and clausal constructions appearing in the corpus. The gold standard provides an indication of a theoretical upper bound on how well the model might aspire to perform, given sufficient data.

The table on the left side of Figure 8.6 lists the relevant comprehension-based measures. Of these, the token and coverage scores, defined in Chapter 4 in terms of the constructional, form and meaning domains, provide the most direct measures of learning performance; their improvement over the course of learning is shown in Figure 8.2.2. All of these show significant improvement between the pre- and post-learning scores, in some cases even approaching gold standard performance. Both form and meaning coverage exhibit a marked increase in the percentage of units (*i.e.*, form or meaning schemas) and relations accounted for by the analysis. (Note that meaning precision is high even initially, indicating that most items in the best-scoring semspec are reliably resolved to some context item.) The combined form-meaning score and overall token score (which

		pre	post	gold
Token (resolved analysis)		45.5	54.6	62.2
Coverage	form (recall$_f$)	53.5	68.0	79.1
	meaning (fscore$_m$)	52.1	58.4	65.2
	precision$_m$	91.7	92.1	96.3
	recall$_m$	37.3	44.3	50.2
	form-meaning	52.8	63.2	72.2
Analysis	|roots|	4.6	3.1	2.3
	|analyses|	1.6	3.1	1.2
	|forms|	3.7	5.6	8.4
	|meanings|	14.5	13.8	12.8
	path density	3.2	9.0	10.1
	value density	41.3	38.6	36.4

	pre	post	gold
length(G)	1711	2246	3054
length(D|G)	726	562	430
length(G|D)	2437	2808	3484

Figure 8.6. Evaluation metrics for learning based on the FT corpus, including comprehension statistics based on token score, form and meaning coverage, and analysis size (left), as well as length measures for the grammar and data (right). All values are shown both pre- and post-learning, along with gold standard values (based on a hand-crafted grammar) for comparison.

includes a component for the constructional analysis) also show consistent upward trends, indicating the learner's gradually improving comprehension, as reflected by its increasingly complete grasp of input utterances and situations.

The table also includes some measures of analysis size. As might be expected, the learned grammar has fewer average roots per analysis, and it accounts for more forms (both words and, crucially, ordering relations among them) than the initial grammar. That is, the learned grammar accounts for more form information with fewer constructions that have greater spans, though it does not achieve quite the level of coverage of the gold standard. The increased number of analyses may reflect the emergence of ambiguity noted earlier. (The lower value for the hand-crafted grammar likely results from deliberate effort on the part of the grammar-writer to produce a minimally redundant grammar.)

The semspec-based measures are a bit harder to interpret, in particular the potentially misleading meaning size metric. As a measure of the number of schemas and role-values accounted for by the analysis, the meaning size might be expected to rise over time, reflecting the learner's increasing capacity for linguistic (and conceptual) complexity. Note, however, that relational constructions perform the crucial function of enforcing semantic bindings (*i.e.*, identity relations) across roles, thereby reducing the number of *independent* meaning schemas. The decrease in the overall mean-

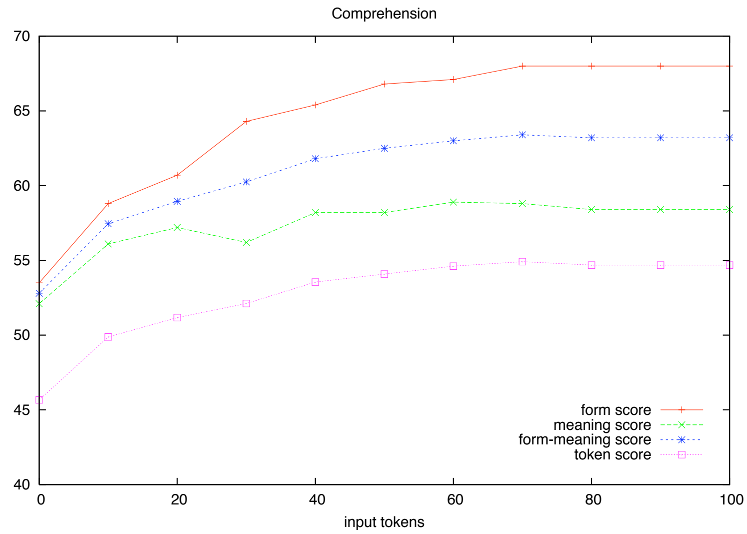

Figure 8.7. Comprehension scores on the FT corpus, including form and meaning coverage as well as the overall token score. All scores display a general trend toward improved comprehension, with performance leveling off around 70 tokens.

ing size thus reflects a drop in the number of (unrelated) meanings as learning progresses. (This drop is relative to the initial lexical grammar, which posits a similar number of schemas but enforces fewer identity bindings among them.) This feature of learning is correlated more directly with the path density value, which measures the average number of (constructional) paths to a particular semspec slot. As shown in the table, this number increases dramatically from the initial lexical grammar to the learned grammar, and is even higher in the gold standard. This metric might be taken as a measure of *coherence*: the more interconnected the meanings in a scene are, the more coherent it is. (The value density measure is less informative, but its increase is presumably attributable to the shrinking number of independent meanings in more advanced grammars.)

The table on the right side of Figure 8.6 shows the changes in the length criteria defined in Chapter 7 for grammars and corpora. As expected, the grammar length $\mathcal{L}(G)$ rises, the data length $\mathcal{L}(D|G)$ falls, and the total description length $\mathcal{L}(G|D)$ rises overall over the course of learning. The absolute numbers themselves are less informative, however, than their progress during training, which provides a window onto the internal dynamics of the learner. Figure 8.8 shows the learner's description length metrics for both external testing (conducted on the held-out test corpus), for comparison, internal testing (conducted on the internal evaluation corpus).

In general, the grammar length is expected to increase over time (as operations add new con-

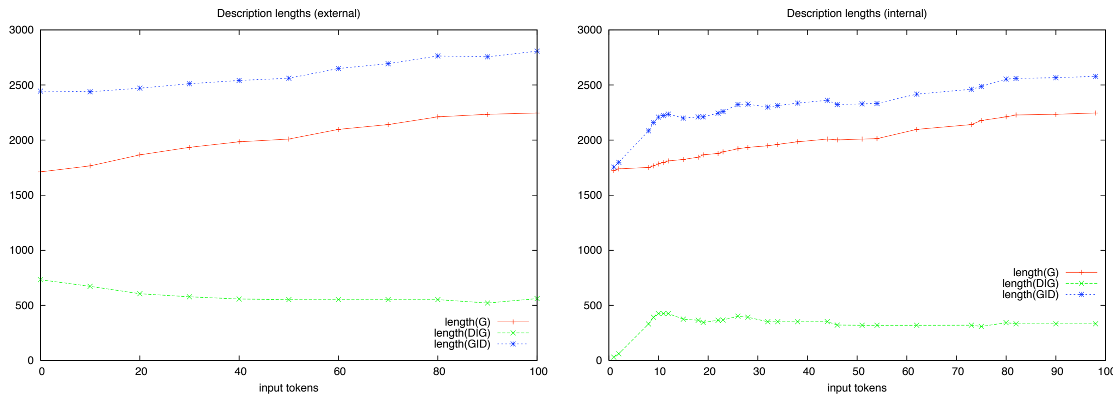

Figure 8.8. Learner performance on length criteria for the FT corpus: both external (left) and internal (right) measures of grammar length, data length and total description length. (The internal test set used by the learner is initially empty, accounting for the initial increase in the data and total description lengths on the right.)

structions), while the data length is expected to decrease (as more useful constructions allowing more compact descriptions of the data are acquired). The external test results exhibit just such behavior, with the data length decreasing during the first half of training before leveling off; this plateau may indicate that the learner has exhausted the instructive potential of the small data set. The total description length, which combines the grammar and data length, also rises gently during training; here, the grammar length is the dominating factor, though this effect could be ameliorated by adjusting the grammar weighting factor.

The length criteria as measured on the internal evaluation set (on the right of Figure 8.8 differ slightly from the external test, notably in its initial climb for both data and total description length and the more fitfull progress thereafter. (The grammar length is the same for both internal and external testing, since it is measured on the same grammar.) Recall, however, that the model's data length is deliberately chosen as a tractable, cognitively plausible subset of encountered tokens, based on recency and salience. The initial increase in data length thus reflects the growing number of tokens in the internal test set; thereafter the data length also improves (decreases) slightly before flattening out. For current purposes, the overall performance trends demonstrate that the tradeoff between grammar and data length (as measured on the internal evaluation set) has predictive value with respect to the external test set, despite the adaptations to standard optimization procedures. Likewise, the plateau encountered during learning may serve as a stopping criterion to prevent overfitting to the data.

8.3 Learning English motion constructions

To test the model on a larger and more diverse range of input, a second experiment was conducted using a subcorpus CSM of the Nomi data, consisting of 200 tokens based on caused- and self-motion predicates. The corpus features a wide range of high-frequency verbs appearing in a variety of syntactic constructions; these include both concrete action verbs (*e.g.*, *fall* and *throw*, as in the FT study, as well as *push*, *drop*, *lie*) and several verbs whose meanings range more freely across motion-based scenes (*e.g.*, *put*, *get*, *come* and *go*). Preliminary results are summarized and discussed below.

8.3.1 Results

Qualitative. The general course of learning resembled that of the previous experiment: many concrete word combinations were first acquired through mapping operations; later merge operations produced constructions exhibiting limited variability, eventually leading to item-specific constructions with more general arguments. The larger corpus size of this experiment was a mixed blessing, however: while a greater variety of constructions were learned, a few challenges and limitations were also apparent. Results of learning are highlighted below.

Concrete word combinations. These included motion-location pairs (`Come-Here`, `Lie-Down`, `Pop-Up`, `Pull-Down`, `Go-Outside`, `Coming-Home`); actor-action pairs (`You-Put`, `Goldie-Fell`); action-object pairs (`Drop-It`, `Put-Blanket`, `Pour-Coffee`); and, interestingly, pairs expressing trajector-landmark relations (`In-Boat`, `On-Floor`).

Variable word combinations. As with the FT experiment, constructions allowing minimal variability were acquired based on collocational overlap. Here, however, the merged constituent categories were motivated by not just object names but also lexical items referring to actions, directions and other relational meanings. Examples of these include `GoingComingE-Home`, `You-Get-DownUpE` (where *get* is used in its self-motion sense), and `Put-Goldie-ThroughDownE`.

Constituency. Sequences with internal constituents were acquired based on both concrete construction pairs like `(You, Lie-Down)` and `(Down, On-Floor)` and more variable constructional categories like `(AnimateL, GoingComingE-Home)`.

Islands and archipelagos. While verbal and other item-specific constructions initially developed independently, in several cases parallel trajectories across syntactically and semantically similar constructions eventually led to the proposal of merge operations like the ones below:

```
MERGE (You−FallL−Down, You−Lie−Down) ===> You−FallLLieE−Down
MERGE (Get−Down, FallL−Down)         ===> MoveTranslationalL−Down
MERGE (Drop−It−TrajectorLandmarkL , Put−This−Through)
                                     ===> DropPutE−EntityL−TrajectorLandmarkL
```

The first two merges above involve scenes of directed self-motion, motivated by phrases like *get/fall/lie down*; the third one recognizes similarities between two scenes of caused motion. The resulting merged relational constructions exhibit increasing variability in their constituents, and may represent steps toward more general argument structure constructions. The `DropPutE-EntityL-TrajectorLandmark` construction learned from the third merge, for example, is still relatively verb-specific, but it nevertheless allows respectable analyses of tokens like the one below:

```
CSM.23: do you put this on the floor
        Do, You, DropPutE−EntityL−TrajectorLandmarkL [Put, This,
                       On−Floor [ On, Floor ] {skipped: TheU}   44.1 (F 48.3,M 40.0)
```

This snapshot suggests how even fairly early on, individual verbs — far from being each one ‘an island, entire of itself’ — may begin to develop links to their neighbors in a greater syntactic and semantic archipelago.

Proto-grammar: A few examples are suggestive of the system’s potential for learning grammatical constructions. The concrete construction `Your-Head` was acquired based on the input utterance *put your head down*, using the “map unmatched” heuristic after a semi-general `Put-EntityL-Down` construction had matched the rest of the utterance:

```
MAP (YourU, Head)           ===>      Your−Head[dG=41.0 dD=−0.179 dGD=20.4]
     MAP (y: YourU, h: Head): y meets h ==== h.m.owner <−−> DS.addressee
```

The context included the constraint that the head in question belongs to Nomi (who is also the discourse addressee), leading to the constructional-discourse constraint noted above as `h.m.owner <--> DS.addressee`. Similarly, the utterance *don’t fall down*, after a partial match based on the `Fall-DownL` construction, led to the proposal of a `Dont-Fall-DownL` construction with two discourse-based constraints, specifying that the speech act type is ’NIMP’ (negative imperative) and binding the addressee to the mover of the falling motion:

```
MAP (DontU, Fall−DownL)          ===>      Dont−Fall−DownL[dG=89.0 dD=−15.2 dGD=12.8]
     MAP (d: DontU, fd: Fall−DownL): d meets fd ==== DS.speechActType <−− NIMP,
                                                     DS.addressee <−−> fd.m.mover
```

These proto-grammatical constructions demonstrate both how unknown words (*i.e.*, ones not present in the intial grammar), even ones without an obvious ostensive referent, may still be learned by the model as parts of larger relational constructions. Though constructions of this kind were too few and far to lead to much generalization in the current experiment, their inclusion is a useful indicator of the model’s representative potential.

Quantitative. Figure 8.9 and Figure 8.10 show the evaluation metrics for the CSM corpus, again in terms of both comprehension and description length. Performance improves for both over the course of learning, with greater coverage, fewer roots, higher path density, and decreased data description length. While these trends are similar to those observed in the FT case, the amount of improvement is generally much smaller, and the gap between the achieved performance and the performance of the gold standard grammar is significantly greater.

		pre	post	gold
Token (resolved analysis)		40.0	46.9	52.3
Coverage	form (recall$_f$)	48.3	53.9	71.6
	meaning (fscore$_m$)	54.2	58.5	71.2
	precision$_m$	86.2	90.6	92.7
	recall$_m$	39.5	43.2	57.8
	form-meaning	51.3	56.2	71.4
Analysis	|roots|	7.0	6.3	3.4
	|analyses|	1.5	2.0	1.7
	|forms|	5.4	6.0	10.8
	|meanings|	19.4	18.9	18.3
	path density	3.1	3.8	8.9
	value density	29.7	29.3	28.0

	pre	post	gold
length(G)	1934	2612	3433
length(D|G)	892	828	637
length(G|D)	2826	3440	4070

Figure 8.9. Evaluation metrics for learning based on the CSM corpus, for both comprehension (left) and description length (right), with gold standard values for comparison.

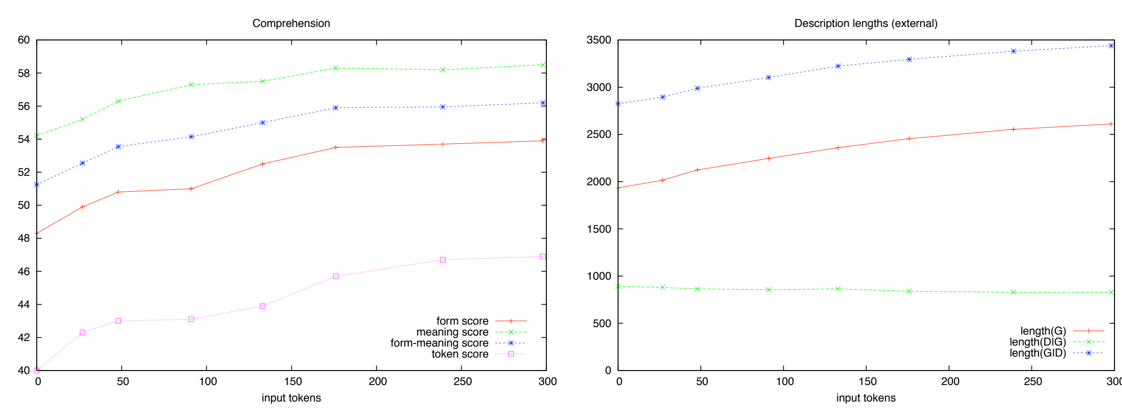

Figure 8.10. Learner performance on the CSM corpus: progress in both comprehension (left) and description length (right), measured on the external test set.

The apparent ceiling in learning improvement may be due in part to the increased complexity and noise of the CSM corpus; even the hand-crafted grammar does not perform especially well here. Learning may be further impaired by the data's relative sparsity: tokens involving similar lexical items are spread through the corpus, often not within a restricted enough window to trigger structure alignment and thereby lead to merge operations. Moreover, there may simply not be enough argument diversity in the data to justify the appropriate generalizations. It is also possible that the model simply requires more training; the metrics in Figure 8.10 may not have yet reached a plateau, and continued iterations through the corpus could yield further improvements. Finally, the model was run using relatively conservative parameters, primarily to restrict the search space to a tractable size; additional experimentation with more permissive assumptions could also lead to improved results.

8.3.2 Limitations and extensions

Both of the experiments described were geared toward the acquisition of relational, English motion-based constructions, primarily those that may shed light on the acquisition of argument structure. In this section I briefly consider some linguistic phenomena not directly addressed by the experiments and discuss whether and to what extent they result from inherent limitations of the model.

Grammatical morphemes. The test corpus included minimal morphological segmentation information for some common affixes (*e.g.*, *-'s*), which in the current experiments were effectively treated as separate unknown words, and not included as candidates for mapping or splitting operations. It seems likely, however, that the model could be extended to learn simple morphological constructions, given appropriately annotated data. As shown in the earlier `Your-Head` example, the relevant representative power is sufficient to acquire a possessive construction based on *-'s*. The corpus includes the utterance *on Daddy -'s foot*, which could easily support the acquisition of a `Daddy-S-Foot` construction based on a meaning binding between the father and his foot, *i.e.*, `f:Foot.m.owner <-- d:Daddy.m`, and ultimately generalized to a `X-S-Y` construction, given appropriate data. Segmented suffixes (*-ed, -ing*)could be similarly mapped to constrain the temporality or aspect associated with an utterance.

Reference. A variety of referential function words (determiners, quantifiers, pronouns) were also not targeted for learning in this experiment. Some of these are good candidates for being learnable in much the same way the earlier discourse-based constructions were. The determiner *the*, for example, might be learned as part of a relational construction specifying the object of joint attention in the current discourse space (`DS.joint-attn`). In general, however, the discourse- and reference-based representations simplified for the current experiments would need to be extended to accommodate more complex referential phenomena.

Auxiliaries. The Nomi corpus includes many instances of verbal auxiliaries, appearing in both standard declarative sentences (*he didnt go anywhere*) and inverted questions (*would you like to get down now?*. Assuming the predication schema can be extended to accommodate the relevant meanings (*e.g.*, tense, modality, aspect), learning such constructions would require bindings across the verb, auxiliary and overall predication — especially in the discontiguous case. The higher cost of encoding all the requisite constraints might make these relatively more difficult to learn.

8.4 Summary

This chapter has presented simple experimental results designed to demonstrate the feasibility of the general approach to language learning taken here, as well as to illuminate some of the inner workings of the model. While further experimentation with larger and richer 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 of examples. Learned constructions also permit a limited degree of generalization, thus allowing for increasingly complete comprehension of new utterances and fulfilling the main goal of the learning model.

Despite many disclaimers that apply to the various simplifying assumptions made here, the model shows both qualitative and quantitative promise for scaling to more realistic linguistic phenomena tackled under more data-rich situations. The emphasis has been on developing an overall framework that is representationally adequate for capturing the relational, compositional nature of grammatical knowledge, and that explicitly integrates concrete theories of language acquisition and use. At this stage, the experimental evidence should be taken primarily as a proof of concept, showing that the model's proposed usage-based learning operations and simplicity-driven evaluation heuristics can, when applied to input that reasonably approximates the child's learning environment, support the acquisition of improved grammars that in turn support improved comprehension. Much work remains on multiple fronts to narrow the gap between this proof of concept and the messy (but increasingly well-documented) realities of child language learning and use; issues involved in such an undertaking, along with many other potential routes forward, are addressed in the chapter ahead.

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.

Bibliography

ABBOT-SMITH, KIRSTEN, ELENA LIEVEN & MICHAEL TOMASELLO. 2001. What children do and do not do with ungrammatical word orders. *Cognitive Development* 16, 1–14.

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

AKHTAR, NAMEERA. 1999. Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language* 26(2), 339–356.

ALISHAHI, AFRA. 2008. *A Probabilistic Model of Early Argument Structure Acquisition*. University of Toronto dissertation.

—— & SUZANNE STEVENSON. 2008. A probabilistic model of early argument structure acquisition. *Cognitive Science* 32, 789–834.

ALLEN, JAMES F. 1984. Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(1), 832–843.

ANDERSON, JOHN R. 1993. *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

ANGLUIN, DANA. 1988. Queries and concept learning. *Machine Learning* 2, 319–342.

—— & CARL H. SMITH. 1983. Inductive inference: Theory and methods. *ACM Computing Surveys* 15(3), 237–269.

ARBIB, MICHAEL A. 2006. The mirror system hypothesis on the linkage of action and languages. In *Action to language via the mirror neuron system*, ed. by Michael A. Arbib, chapter 1. New York: Cambridge University Press.

—— & GIACOMO RIZZOLATTI. 1997. Neural expectations: a possible evoluationary parth from manual skills to language. *Communications and Cognition* 29, 393–424.

ASLIN, RICHARD N., JENNY R. SAFFRAN & ELISSA L. NEWPORT. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9, 321–324.

BAILEY, DAVID R. 1997. *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. Computer Science Division, University of California at Berkeley dissertation.

——, JEROME A. FELDMAN, SRINI NARAYANAN & GEORGE LAKOFF. 1997. Modeling embodied lexical development. In *Proc. 19th Cognitive Science Society Conference*.

BAKER, CARL L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533–581.

BALDWIN, DARE A. 1993. Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language* 20, 395–418.

BARSALOU, LAWRENCE W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–660.

BATES, ELIZABETH. 1976. *Language and Context: The Acquisition of Pragmatics*. Academic Press.

—— & BRIAN MACWHINNEY. 1987. Competition, variation, and language learning. In (MacWhinney 1987), 157–193.

BERGEN, BENJAMIN. 2004. The psychological reality of phonaesthemes. *Language* 80(2), 290–311.

BERGEN, BENJAMIN K. & NANCY CHANG. 2005. Simulation-based language understanding in Embodied Construction Grammar. In *Construction Grammars: Cognitive and Cross-language dimensions*, ed. by Jan-Ola Östman & Mirjam Fried. John Benjamins.

BERLIN, B. & P. KAY. 1969. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

BLOOM, LOIS. 1973. *One word at a time: The use of single word utterances before syntax*. The Hague: Mouton & Co.

BLOOM, PAUL. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

BLOOMFIELD, LEONARD. 1931. *Review of J. Ries, Was ist ein Satz?*. In (Hockett 1970).

——. 1943. *Meaning*. In (Hockett 1970).

BOAS, HANS CHRISTIAN. 2003. *Resultative constructions in English and German*. Stanford, CA: CSLI Publications.

BOWERMAN, MELISSA. 1976. Semantic factors in the acquisition of word use and sentence construction. In *Directions in normal and deficient language*, ed. by D. Morehead & A. Morehead. Baltimore, Md.: University Park Press.

——. 1988. The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In *Explaining language universals*, ed. by J. Hawkins, chapter 4. London: Blackwell.

BRAINE, MARTIN D.S. 1963. The ontogeny of English phrase structure: The first phase. *Language* 39(1), 1–14.

——. 1976. *Children's first word combinations*, volume 41 of *Monographs of the Society for Research in Child Development*. University of Chicago Press.

BRENT, MICHAEL R. 1999. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning* 34(1-3), 71–105.

—— & TIMOTHY A. CARTWRIGHT. 1996. Lexical categorization: Fitting template grammars by incremental MDL optimization. In *Grammatical Inference: Learning Syntax from Sentences*, volume 1147 of *Lecture Notes in Computer Science*, 84–94. Springer-Verlag.

BRESNAN, JOAN. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

BROOKS, PATRICIA J. & MICHAEL TOMASELLO. 1999. Young children learn to produce passives with nonce verbs. *Developmental Psychology* 35, 29–44.

BROWN, ROGER W. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

BRYANT, JOHN. 2003. Constructional analysis. Master's thesis, University of California at Berkeley.

——. 2008. *Best-Fit Constructional Analysis*. Computer Science Division, University of California at Berkeley dissertation.

BUCCINO, G., F. BINKOFSKI, G.R. FINK, L. FADIGA, L. FOGASSI, V. GALLESE, R.J. SEITZ, K. ZILLES, G. RIZZOLATTI & H.-J. FREUND. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience* 13, 404–404.

BUDWIG, NANCY. 1995. *A Developmental-Functionalist Approach to Child Language*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

BYBEE, JOAN. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam, Philadelphia: John Benjamins.

——. 2001. *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.

——. 2006. From usage to grammar: The mind's response to repetition. *Language* 82, 711.

CANGELOSI, ANGELO. 2005. The emergence of language: Neural and adaptive agent models. An introduction. *Connection Science* 17(3-4), 185–190.

——. 2006. The grounding and sharing of symbols. *Pragmatics and Cognition* 14(2), 275–285.

CAREY, SUE. 1978. The child as word-learner. In *Linguistic theory and psychological reality*, ed. by M. Halle, J. Bresnana, & G.A. Miller. Cambridge, MA: MIT Press.

CHANG, NANCY, JEROME FELDMAN, ROBERT PORZEL & KEITH SANDERS. 2002a. Scaling cognitive linguistics: Formalisms for language understanding. In *Proc. 1st International Workshop on Scalable Natural Language Understanding*, Heidelberg, Germany.

——, DANIEL GILDEA & SRINI NARAYANAN. 1998. A dynamic model of aspectual composition. In *Proc. 20th Cognitive Science Society Conference*, 226–231, Madison, Wisconsin.

—— & EVA MOK. 1998. A structured context model for grammar learning. In *Proc. International Joint Conference on Neural Networks*, 226–231, Vancouver, BC.

——, SRINI NARAYANAN & MIRIAM R.L. PETRUCK. 2002b. Putting frames in perspective. In *Proc. 19th International Conference on Computational Linguistics*, Taipei, Taiwan.

CHATER, NICK. 2004. What can be learned from positive data? Insights from an 'ideal learner'. *Journal of Child Language* 31, 915–918.

—— & CHRIS MANNING. 2006. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences* 10(7), 335–344.

—— & PAUL VITÁNYI. 2007. 'ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psycholgy* 51(3), 135–163.

—— & PAUL M.B. VITÁNYI. 2003. Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences* 7, 19–22.

CHOI, SOONJA & MELISSA BOWERMAN. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41, 81–121.

CHOMSKY, NOAM. 1955. Semantic considerations in grammar. *Georgetown Monograph Series on Language and Linguistics* (8), 141–158.

——. 1957. *Syntactic Structures*. The Hague: Mouton.

——. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Press.

CLARK, ALEXANDER. 2001. *Unsupervised Language Acquisition: Theory and Practice*. University of Sussex dissertation.

CLARK, EVE V. 1993. *The Lexicon in Acquisition*. Cambridge, UK: Cambridge University Press.

——. 2003. *First Language Acquisition*. Cambridge, UK: Cambridge University Press.

CLARK, HERBERT H. (ed.). 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.

CROFT, WILLIAM. 1990. A conceptual framework for grammatical categories (or, a taxonomy of propositional acts). *Journal of Semantics* 7, 245–79.

——. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.

——. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

CULICOVER, PETER W. & RAY JACKENDOFF. 1999. The view from the periphery: The English correlative conditional. *Linguistic Inquiry* 28(2), 195–217.

CULY, CHRISTOPHER. 1985. The complexity of the vocabulary of Bambara. *Linguistics and Philosophy* 8, 345–351.

DALRYMPLE, MARY. 2001. *Lexical Functional Grammar*. New York: Academic Press.

DE BEUELE, JOACHIM. 2008. The emergence of compositionality, hierarchy and recursion in peer-to-peer interactions. In *Proceedings of the 7th International Conference on the Evolution of Language*, ed. by A. D. M. Smith, K. Smith, & R. Ferrer i Cancho, 75–82. World Scientific.

DE BEULE, JOACHIM & BENJAMIN K. BERGEN. 2006. On the emergence of compositionality. In *Proceedings of the 6th International Conference on the Evolution of Language*, 35–42.

—— & LUC STEELS. 2005. Hierarchy in fluid construction grammar. In *Proceedings of KI-2005*, ed. by Furbach U., number 3698 in Lecture Notes in AI, 1–15, Berlin. Springer-Verlag.

DE MARCKEN, CARL. 1995. *Unsupervised Language Acquisition*. Massachusetts Institute of Technology dissertation.

DE SAUSSURE, FERDINAND. 1916. *Course de linguistique générale*. Paris: Payot.

DEACON, TERRENCE W. 1997. *The symbolic species: The co-evolution of language and the brain*. New York: W.W. Norton.

DEJONG, GERALD & RAYMOND J. MOONEY. 1986. Explanation-based learning: An alternative view. *Machine Learning* 1(2), 145–176.

DODGE, ELLEN. In Preparation. *Conceptual and Constructional Composition*. University of California at Berkeley dissertation.

—— & GEORGE LAKOFF. 2005. Image schemas: From linguistic analysis to neural grounding. In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, ed. by Beate Hampe, 57–91. Berlin: Mouton de Gruyter.

DOMINEY, PETER FORD & JEAN-DAVID BOUCHER. 2005. Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence* 167(1-2), 31–61.

EDELMAN, GERALD M. 2007. *Bright Air, Brilliant Fire*. Basic Books.

EDELMAN, SHIMON. 2004. Bridging language with the rest of cognition: computational, algorithmic and neurobiological issues and methods. In *Proc. of the Ithaca workshop on Empirical Methods in Cognitive Linguistics*.

——, ZACH SOLAN, DAVID HORN & EYTAN RUPPIN. 2004. Bridging computational, formal and psycholinguistic approaches to language. In *Proc. 26th Cognitive Science Society Conference*.

ELMAN, JEFFREY L., KIM PLUNKETT, DOMENICO PARISI, ANNETTE KARMILOFF-SMITH, MARK H. JOHNSON & ELIZABETH A. BATES. 1996. *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

EVANS, C.R. & A.D.J. ROBERTSON (eds.). 1968. *Key Papers: Cybernetics*. Baltimore, MD: University Park Press.

FALKENHAINER, BRIAN, KENNETH D. FORBUS & DEDRE GENTNER. 1989. The Structure-Mapping Engine: Algorithm and examples. *Aritificial Intelligence* 20(41). 1-63.

FAUCONNIER, GILLES. 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, Mass. and London: MIT Press/Bradford.

——. 1987. *Mappings in Thought and Language*. Cambridge, UK: Cambridge University Press.

—— & MARK TURNER. 2003. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

FELDMAN, JEROME A. 1972. Some decidability results on grammatical inference and complexity. *Information and Control* 20, 244–262.

——. 1982. Dynamic connections in neural networks. *Biological Cybernetics* 46, 27–39.

——. 2006. *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.

—— & DANA H. BALLARD. 1982. Connectionist models and their properties. *Cognitive Science* 6, 205–254.

——, ELLEN DODGE & JOHN BRYANT. To appear. A neural theory of language and embodied construction grammar. In *The Oxford Handbook of Linguistic Analysis*, ed. by B. Heine & H. Narrog. Oxford University Press.

——, GEORGE LAKOFF, DAVID BAILEY, SRINI NARAYANAN, TERRY REGIER & ANDREAS STOLCKE. 1996. L_0—the first five years of an automated language acquisition project. *AI Review* 10, 103–129.

FILLMORE, CHARLES. 1968. The case for case. In *Universals in Linguistic Theory*, ed. by E. Bach & R. Harms. New York: Holt, Rinehart, and Winston.

——, PAUL KAY & MARY C. O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501–538.

FILLMORE, CHARLES J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, ed. by Linguistic Society of Korea, 111–38. Seoul: Hanshin.

GAHL, SUSANNE & SUSAN GARNSEY. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80, 748–775.

——. 2006. Knowledge of grammar includes knowledge of syntactic probabilities. *Language* 82, 405–410.

GALLESE, V., L. FADIGA, L. FOGASSI & G. RIZZOLATTI. 1996. Action recognition in the premotor cortex. *Brain* 119, 593–609.

—— & G. LAKOFF. 2005. The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 22, 455–479.

GENTNER, DEDRE. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2).

—— & ARTHUR B. MARKMAN. 1997. Structure mapping in analogy and similarity. *American Psychologist* 52, 46–56.

—— & LAURA L. NAMY. 2006. Analogical processes in language learning. *Current Directions in Psychological Science* 15(6), 297–301.

GLEITMAN, LILA R. & JANE GILLETTE. 1995. The role of syntax in verb learning. In *The Handbook of Child Language*, ed. by P. Fletcher & B. MacWhinney. Cambridge, MA: Basil Blackwell.

GLENBERG, ARTHUR M. & MICHAEL P. KASCHAK. 2002. Grounding language in action. *Psychonomic Bulletin & Review* 9, 558–565.

—— & DAVID A. ROBERTSON. 1999. Indexical understanding of instructions. *Discourse Processes* 28, 1–26.

GOLD, E. MARK. 1967. Language identification in the limit. *Information and Control* 16, 447–474.

GOLDBERG, ADELE E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

——, DEVIN M. CASENHISER & NITYA SETHURAMAN. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15, 289–316.

—— & RAY JACKENDOFF. 2004. The English resultative as a family of constructions. *Language* 80(3), 532–568.

GOLDSMITH, JOHN A. 2002. Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27(27), 153–198.

GÓMEZ, REBECCA & LOUANN GERKEN. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4(5), 178–186.

—— & LOUANN GERKEN. 2002. Variability and detection of invariant structure. *Psychological Science* 13, 431–436.

GOPNIK, ALISON. 1982. Words and plans: Early language and the development of intelligent action. *Journal of Child Language* 9, 303–318.

GRADY, JOSEPH E. 1997. *Foundations of Meaning: Primary Metaphors and Primary Scenes*. University of California at Berkeley dissertation.

GRICE, H.P. 1975. Logic and conversation. In *Syntax and Semantics, 3: Speech Acts*, ed. by P. Cole & J.L. Morgan. New York: Academic Press.

HAUK, O., I. JOHNSRUDE & F. PULVERMÜLLER. 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41(2), 301–307.

HAUSER, M., N. CHOMSKY & W.T. FITCH. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598), 1569–1579.

HINTON, L., J. NICHOLS & J. OHALA (eds.). 1996. *Sound Symbolism*. Cambridge University Press.

HIRSH-PASEK, KATHY & ROBERT MICHNICK GOLINKOFF. 1996. *The Origins of Grammar: Evidence from Early Language Comprehension*. Cambridge, MA; London, England: MIT Press.

HOCKETT, C.F. (ed.). 1970. *A Leondard Bloomfield Anthology*. Bloomington, IN: Indiana University Press.

HOPPER, PAUL J. & ELIZABETH CLOSS TRAUGOTT. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.

HORNING, JAMES JAY. 1969. A study of grammatical inference. Technical Report CS-139, Stanford University Computer Science Department.

HUCK, GEOFFREY J. & JOHN A. GOLDSMITH. 1995. *Ideology and linguistic theory: Noam Chomsky and the deep structure debates*. London, New York: Routledge.

HUDSON KAM, CARLA L. & ELISSA L. NEWPORT. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1, 151–195.

ISRAEL, MICHAEL. To appear. How children get constructions. In *Pragmatics in Construction Grammar and Frame Semantics*, ed. by Mirjam Fried & Jan-Ola Östman. Johns Benjamins.

JACKENDOFF, RAY. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

JELINEK, FRED. 1988. Applying information theoretic methods: Evaluation of grammar quality. Presentation at Workshop on Evaluation of NLP Systems, Wayne, PA.

JOHNSON, CHRISTOPHER R. 1999. *Constructional grounding: The role of interpretational overlap in lexical and constructional acquisition*. University of California at Berkeley dissertation.

JOHNSON, MARK. 1987. *The Body in the Mind*. University of Chicago Press.

JURAFSKY, DANIEL & JAMES H. MARTIN. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

JUSCZYK, PETER W. 1997. Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science* 6, 170–174.

KASCHAK, MICHAEL P. & ARTHUR M. GLENBERG. 2000. Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language* 43, 508–529.

KAY, PAUL & CHARLES FILLMORE. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* 75(1), 1–33.

—— & TERRY REGIER. 2006. Language, thought and color: Recent developments. *TRENDS in Cognitive Science* 10(2), 51–54.

KLEIN, DAN & CHRIS MANNING. 2004. Corpus-based induction of syntactic structure: Models of dependency and constituency. In *Proceedings of the Association for Computational Linguistics*.

LAKOFF, GEORGE. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

—— & MARK JOHNSON. 1980. *Metaphors We Live By*. University of Chicago Press.

LAMBRECHT, KNUD. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.

LANDAU, BARBARA, LINDA B. SMITH & SUSAN S. JONES. 1988. Object shape, object function, and object name. *Journal of Memory and Language* 38, 1–27.

LANGACKER, RONALD W. 1987. *Foundations of Cognitive Grammar, Vol. 1*. Stanford University Press.

——. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Cognitive Linguistics Research. Berlin and New York: Mouton de Gruyter.

LEE, LILLIAN. 1996. Learning of context-free languages: A survey of the literature. Technical Report TR-12-96, Harvard University.

LI, MING & PAUL M. B. VITÁNYI. 1997. *An Introduction to Kolmogorov Complexity and Its Applications*. Berlin: Springer-Verlag.

LIEVEN, ELENA, HEIKE BEHRENS, JENNIFER SPEARERS & MICHAEL TOMASELLO. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30, 333–370.

LIEVEN, ELENA V.M., JULIAN M. PINE & GILLIAN BALDWIN. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24, 187–219.

MACKAY, DAVID J.C. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

MACWHINNEY, BRIAN (ed.). 1987. *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

——. 1998. Models of the emergence of language. *Annual Review of Psychology* 49, 199–227.

——. 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31, 883–914.

——. 2005. The emergence of grammar from persective. In *The Grounding of Cognition: The role of Perception and Action in Memory, Language, and Thinking*, ed. by Diane Pecher & Rolf A. Zwaan. Mahwah, NJ: Lawrence Erlbaum Associates.

MANDLER, JEAN M. 1988. How to build a baby: On the development of an accessible representational system. *Cognitive Development* 3, 113–136.

——. 1992. How to build a baby: II. Conceptual primitives. *Psychological Review* 99, 587–604.

MARATSOS, MICHAEL P. & MARY ANNE CHALKLEY. 1980. The internal language of children's syntax. In *Children's language*, ed. by K. Nelson, volume 2, 127–213. New York: Gardner Press.

MARCUS, GARY F. 1993. Negative evidence in language acquisition. *Cognition* 46, 53–85.

MARKMAN, ARTHUR B. 1999. *Knowledge Representation*. Lawrence Erlbaum Associates.

MARKMAN, ELLEN M. 1989. *Categorization and naming in children*. Cambridge, MA: MIT Press.

MARR, DAVID. 1982. *Vision: A Computational Investigation into Human Representation and Processing of Visual Information*. San Francisco: Freeman and Co.

MATLOCK, TEENIE. 2003. Fictive motion as cognitive simulation. *Memory & Cognition* 32, 1328–1400.

MAY, SHIRA, ROSANNA MUCETTI, DAN SLOBIN & GAIL SOLOMON. 1996. Coding of motion events in child-adult discourse. Unpublished annotation manual.

MELTZOFF, ANDREW N. 1995. Understanding the intentions of others: Re-enactment of intended acts by eighteen-month-old children. *Developmental Psychology* 31, 838–850.

MICHAELIS, LAURA. 1994. A case of constructional polysemy in Latin. *Studies in Language* 18, 23–48.

MINSKY, MARVIN. 1974. A framework for representing knowledge. MIT-AI Laboratory Memo 306.

MOK, EVA. 2008a. *Contextual Bootstrapping for Grammar Learning*. Computer Science Division, University of California at Berkeley dissertation.

——. 2008b. Extended annotation scheme for CHILDES data. Unpublished annotation manual.

——, JOHN BRYANT & JEROME FELDMAN. 2004. Scaling understanding up to mental spaces. In *HLT-NAACL 2004 Workshop: 2nd Workshop on Scalable Natural Language Understanding*, ed. by Robert Porzel, 41–48, Boston, Massachusetts, USA. Association for Computational Linguistics.

MUGGLETON, STEPHEN & LUC DE RAEDT. 1994. Inductive logic programming: Theory and methods. *Journal of Logic Programming* 19/20, 629–679.

NAMY, LAURA L. 2001. What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy* 2, 73–86.

NARAYANAN, SRINI. 1997a. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. Computer Science Division, University of California at Berkeley dissertation.

——. 1997b. Talking the talk *is* like walking the walk: A computational model of verbal aspect. In *Proc. 19th Cognitive Science Society Conference*.

NELSON, KATHARINE. 1996. *Language in Cognitive Development: The Emergence of the Mediated Mind*. Cambridge, U.K.: Cambridge University Press.

OATES, TIM, ZACHARY EYLER-WALKER & PAUL R. COHEN. 1999. Syntax to learn semantics: An experiment in language acquisition with a mobile robot. Technical report, University of Massachusetts Computer Science Department Technical Report 99-35.

OMOHUNDRO, STEPHEN. 1992. Best-first model merging for dynamic learning and recognition. Technical Report TR-92-004, International Computer Science Institute, Berkeley, CA.

ONNIS, LUCA, MATTHEW ROBERTS & NICK CHATER. 2002. Simplicity: A cure for overregularizations in language acquisition? In *Proc. 24th Cognitive Science Society Conference*, 720–725.

PEIRCE, CHARLES SANDERS. 1976. The new elements of mathematics by charles s. peirce. In *Mathematical Philosophy*, ed. by Carolyn Eisele, volume 4. The Hague, The Netherlands: Mouton Publishers.

PEREIRA, FERNANDO. 2000. Formal grammar and information theory: Together again. *Philosophical Transactions of the Royal Society* 358, 1239–1253.

PERFORS, AMY. 2008. *Learnability, representation, and language: A Bayesian approach*. Massachusetts Institute of Technology dissertation.

——, JOSHUA B. TENENBAUM & TERRY REGIER. 2006. Poverty of the stimulus? A rational approach. In *Proc. 28nd Cognitive Science Society Conference*, volume 28.

PETERS, ANN M. 1985. *Language Segmentation: Operating Principles for the Perception and Analysis of Language*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

PINE, JULIAN M. & ELENA V. LIEVEN. 1993. Reanalysing rote-learned phrases: Individual differences in the transition to multiword speech. *Journal of Child Language* 20, 551–571.

PINKER, STEVEN. 1979. Formal models of language learning. *Cognition* 7, 217–283.

——. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Bradford Books.

——, DAVID S. LEBEAUX & LOREN ANN FROST. 1987. Productivity and constraints in the acquisition of the passive. *Cognition* 26, 195–267.

POLLARD, CARL & IVAN A. SAG. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

PULVERMÜLLER, F., M. HAERLE & F. HUMMEL. 2002. Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language* 78, 143–168.

QUINE, WILLIAM VAN ORMAN. 1960. *Word and Object*. Cambridge, MA: MIT Press.

REGIER, TERRY. 1996. *The Human Semantic Potential*. Cambridge, MA: MIT Press.

——. 1998. Reduplication and the arbitrariness of the sign. In *Proc. 20th Cognitive Science Society Conference*, volume 22, 182–187.

REISIG, WOLFGANG. 1985. *Petri Nets: An Introduction*. Berlin: Springer–Verlag.

RICHARDSON, DANIEL C., MICHAEL J. SPIVEY, KEN MCRAE & LAWRENCE W. BARSALOU. 2003. Spatial representations activated during real-time comprehension of verbs. *Cognitive Science* 27, 767–780.

RISSANEN, JORMA. 1978. Modeling by shortest data description. *Automatica* 14, 465–471.

——. 1989. *Stochastic complexity and statistical inquiry*. World Scientific.

RIZZOLATTI, GIACOMO & MICHAEL A. ARBIB. 1998. Language within our grasp. *Trends in Neurosciences* 21(5), 181–194.

——, LUCIANO FADIGA, VITTORIO GALLESE & LEONARDO FOGASSI. 1996. Premotor cortex and the recognition ofmotor actions. *Cognitive Brain Research* 3, 131–141.

ROY, DEB. 1999. *Learning from Sights and Sounds: A Computational Model*. MIT Media Laboratory dissertation.

——. 2005. Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence* 167(1-2), 170–205.

RUMELHART, DAVID E., JAMES L. MCCLELLAND & THE PDP RESEARCH GROUP. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge: MIT Press.

RUSSELL, STUART & PETER NORVIG. 1994. *Artificial Intelligence: A Modern Approach*. Prentice Hall. ISBN: 0131038052.

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

SAFFRAN, JENNY R., RICHARD N. ASLIN & ELISSA L. NEWPORT. 1996. Statistical learning by 8-month-old infants. *Science* 274, 1926–192.

SAG, IVAN A., THOMAS WASOW & EMILY M. BENDER. 2003. *Syntactic Theory: A Formal Introduction*. Stanford, CA: CLSI Publications.

SAKAKIBARA, YASUBUMI. 1995. Grammatical inference: An old and new paradigm. In *Proc. 6th Int. Workshop on Algorithmic Learning Theory*, 1–24. Springer-Verlag.

SCHLESINGER, IZCHAK M. 1979. Cognitive and linguistic structure: The case of the instrumental. *Journal of Linguistics* 15, 307–324.

SELFRIDGE, MALLORY. 1986. A computer model of child language learning. *Artificial Intelligence* 29, 171–216.

SHASTRI, LOKENDRA. 1988. *Semantic Networks: An evidential formalization and its connectionist realization*. Los Altos, CA: Morgan Kaufmann.

——. 2001. Biological grounding of recruitment learning and vicinal algorithms in long-term potentiation.

—— & VENKAT AJJANAGADDE. 1993. From simple associations to systematic reasoning. *Behavioral and Brain Sciences* 16(3), 417–494.

——, Dean J. Grannes, Srini Narayanan & Jerome A. Feldman. 1999. A connectionist encoding of parameterized schemas and reactive plans. In *Hybrid Information Processing in Adaptive Autonomous Vehicles*, ed. by G.K. Kraetzschmar & G. Palm, Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag.

Shieber, Stuart M. 1985. Evidence against the non-context-freeness of natural language. *Linguistics and Philosophy* 8, 333–343.

——. 1986. *An Introduction to Unification-Based Approaches to Grammar*. Stanford, CA: CSLI Publications.

Sinha, Steven K. 2008. *Answering Questions about Complex Events*. Computer Science Division, University of California at Berkeley dissertation.

Siskind, Jeffrey Mark. 1997. A computational study of cross-situational techniques for learning word-to-meaning mappings. In *Computational Approaches to Language Acquisition*, ed. by Michael R. Brent, chapter 2. Cambridge, MA: MIT Press.

——. 2000a. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. Technical Report 2000-105, NEC Research Institute.

——. 2000b. Visual event classification via force dynamics. In *Proc. AAAI-2000*, 149–155.

——. 2001. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *Artificial Intelligence Research* 15, 31–90.

Slobin, Dan Isaac. 1985. Crosslinguistic evidence for the language-making capacity. In *Theoretical Issues*, ed. by Dan Isaac Slobin, volume 2 of *The Crosslinguistic Study of Language Acquisition*, chapter 15. Mahwah, New Jersey: Lawrence Erlbaum Associates.

——. 1991. Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics* 1, 7–26.

——. 1997. The origins of grammaticizable notions: Beyond the individual mind. In *Expanding the Contexts*, ed. by Dan Isaac Slobin, volume 5 of *The Crosslinguistic Study of Language Acquisition*, chapter 5. Mahwah, New Jersey; London: Lawrence Erlbaum Associates.

Steels, Luc. 1996. Perceptually grounded meaning creation. In *Proceedings of ICMAS*, ed. by M. Tokoro, Kyoto.

——. 1997. The synthetic modeling of language origins. *Evolution of Communication* 1(1), 1–35.

——. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence* 103, 1–24.

——. 2000. The emergence of grammar in communicating autonomous robotic agents. In *Proc. ECAI 2000*, Amsterdam. IOS Publishing.

——. 2006. Experiments on the emergence of human communication. *Trends in Cognitive Sciences* 10, 347–349.

——, Joachim De Beule & Nicolas Neubauer. 2005. Linking in fluid construction grammar. In *BNAIC-05*, Brussels, Belgium. Royal Flemish Academy for Science and Art.

—— & Frédéric Kaplan. 1998. Situated grounded word semantics. *Proceedings IJCAI-99* 862–867.

—— & Paul Vogt. 1997. Grounding adaptive language games in robotic agents. In *Proceedings of the ECAL-97*, ed. by P. Husbands & I. Harvey, Cambridge, MA. MIT Press.

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

—— & STEPHEN M. OMOHUNDRO. 1994. Inducing probabilistic grammars by Bayesian model merging. In *Grammatical Inference and Applications*, ed. by R. C. Carrasco & J. Oncina, 106–118. Springer-Verlag.

TALMY, LEONARD. 1988. Force dynamics in language and cognition. *Cognitive Science* (12), 49–100.

——. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

TETTAMANTI, M., G. BUCCINO, M.C. SACCUMAN, V. GALLESE, M. DANNA, P. SCIFO, F. FAZIO, G. RIZZOLATTI, S.F. CAPPA & D. PERANI. 2005. Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience* 17, 273–281.

THOMPSON, CYNTHIA A. 1998. *Semantic Lexicon Acquisition for Learning Natural Language Interfaces*. Department of Computer Sciences, University of Texas at Austin dissertation.

—— & RAYMOND J. MOONEY. 1999. Automatic construction of semantic lexicons for learning natural language interfaces. In *Proc. 16th National Conference on Artificial Intelligence*, Orlando, FL.

TOMASELLO, MICHAEL. 1992. *First verbs: A case study of early grammatical development*. Cambridge, UK: Cambridge University Press.

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

——. 1999. *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

——. 2000. First steps in a usage based theory of language acquisition. *Cognitive Linguistics* 11, 61–82.

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

—— & PATRICIA J. BROOKS. 1998. Young children's earliest transitive and intransitive constructions. *Cognitive Linguistics* 9(4), 379–395.

TURING, ALAN. 1948. Intelligent machinery. In (Evans & Robertson 1968).

VALIANT, L.G. 1984. A theory of the learnable. *Communications of the ACM* 27, 1134–1142.

WHARTON, R.M. 1974. Approximate language identification. *Information and Control* 26, 236–255.

WHORF, BENJAMIN LEE. 1956. *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

WITTGENSTEIN, LUDWIG. 1958. *Philosophical Investigations*. New York: Macmilllan Publishing Co.

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

WONNACOTT, ELIZABETH, ELISSA L. NEWPORT & MICHAEL K. TANENHAUS. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology* .

WOODWARD, AMANDA L. & KAREN L. HOYNE. 1999. Infants' learning about words and sounds in relation to objects. *Child Development* 70, 65–77.