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

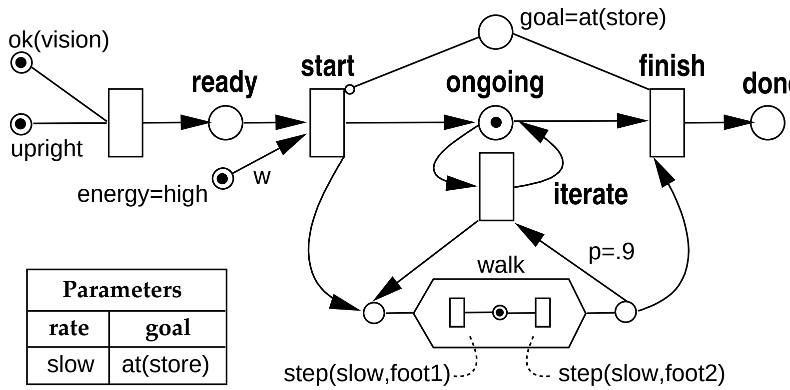

Figure 2.2. X-schema for WALK(TO STORE)

(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.