On look-ahead in language: navigating a multitude of familiar paths*

Shimon Edelman

> Do I understand this sentence? Do I understand it just as I should if I heard it in the course of a narrative? If it were set down in isolation I should say, I don't know what it's about. But all the same I should know how this sentence might perhaps be used; I could myself invent a context for it. (A multitude of familiar paths lead off from these words in every direction.)
>
> — Wittgenstein (1958, §525, p.142)

1 What is the title of this … ?

Language is a rewarding field if you are in the prediction business. A reader who is fluent in English and who knows how academic papers are typically structured will readily come up with several possible guesses as to where the title of this section could have gone, had it not been cut short by the ellipsis. Indeed, in the more natural setting of spoken language, anticipatory processing is a must: performance of machine systems for speech interpretation depends critically on the availability of a good predictive model of how utterances unfold in time (Baker, 1975; Jelinek, 1990; Goodman, 2001), and there is strong evidence that prospective uncertainty affects human sentence processing too (Jurafsky, 2003; Hale, 2006; Levy, 2008).

The human ability to predict where the current utterance is likely to be going is just another adaptation to the general pressure to anticipate the future (Hume, 1748; Dewey, 1910; Craik, 1943), be it in perception, thinking, or action, which is exerted on all cognitive systems by evolution (Dennett, 2003). Look-ahead in language is, however, special in one key respect: language is a medium for communication, and in communication the most interesting (that is, informative) parts of the utterance that the speaker is working through are those that cannot be predicted by the listener ahead of time.

*A preliminary draft of a chapter written for the volume *Prediction in the Brain*, M. Bar, ed., Oxford University Press, forthcoming. SE is with the Dept. of Psychology, Cornell University, Ithaca, NY 14853, USA, and Dept. of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Republic of Korea. File generated on 2009-11-21 00:00.

That certain parts of an utterance or some of the aspects of its structure are unpredictable in a given context does not imply that they can all be entirely novel, that is, never before encountered by the listener in any other context; if too many of them were, communication would have been just as impossible as with completely predictable signals.[1]

In theorizing about how language mediates communication, it is tempting to make the opposite assumption, namely, that both the structures and the parts (lexicon) are fully shared between the interlocutors, with only the assignment of parts to slots in structures being unexpected in the present context and hence informative. This temptation, however, must be firmly resisted; as Quine (1961, p.259) put it, "the narrowly linguistic habits of vocabulary and syntax are imported by each speaker from his unknown past." It is certainly convenient to assume, as the so-called generative tradition in linguistics does,[2] that all humans share an innately specified universal grammar that defines all and only structures that a listener need ever contemplate while processing a speaker's output. Unfortunately, this assumption runs counter to empirical findings even for adults, let alone for infants who are just learning to make sense of the hubbub that surrounds them, and who, in doing so, only gradually overcome the vast mismatch in structural and conceptual knowledge that initially exists between them and their caregivers (Edelman and Waterfall, 2007).

The individual differences among language users, being the rule rather than an exception in language (Chipere, 2001; Dabrowska and Street, 2006), cause structural and conceptual interpretation gaps to open between interlocutors. To understand how linguistic communication is at all possible, we should integrate insights and theories from language development (which lays down the foundation for an individual's linguistic ability), processing (which initially overcomes formidable difficulties; Von Berger, Wulfeck, Bates, and Fink, 1996; Thal and Flores, 2001), and generation (the capacity for which builds up gradually, as the brain matures and assimilated more and more experience; Bloom, 1970; Bates, Thal, Finlay, and Clancy, 1999; Bates and Goodman, 1999; Diessel, 2004). Only such an integrated approach, grounded in abundant and realistic behavioral data (rather than in an intuitive analysis of hand-picked cases), can lead to an understanding both of the nature of the knowledge that is shared by language users and of their idiosyncrasies.

The order of business for the remainder of this chapter is, therefore, as follows. Section 2 proposes a computational framework that seems particularly suitable for the representation and processing of experience data. Section 3 looks at such data in search of cues that may be helping infants learn language reliably and efficiently by turning experience into a kind of grammar. Section 4 then outlines a hypothesis regarding the possible brain mechanisms for acquiring and maintaining linguistic knowledge that fit within the proposed computational framework. Finally, section 5 suggests how the proposed approach may advance the development of new models of language acquisition and processing. As we shall see, prediction — that is, projection of the past experience into the immediate future — figures prominently in all these settings.

2 The structure of sensorimotor experience

The idea of grammar — a formal system that codifies the well-formedness of a class of utterances to the exclusion of others — as the repository of the knowledge of a language arises from the textbook answer to the fundamental question of linguistics: what does it mean to know a language? This answer, however, is only valid if one assumes *a priori* that a person's knowledge of language depends entirely on an ability to tell apart well-formed ("grammatical") sentences from ill-formed ones (Chomsky, 1957).[3] Although this assumption underlies a tremendous amount of work in the linguistic tradition that has been termed formalist, it is not the only game in town: there is a complementary, functionalist, view, which focuses on language *use* (Newmeyer, 1998). From the functionalist standpoint, to know a language means, roughly, to be able to conceptualize what you hear and to be ready to prove that you do by generating an appropriate reply or action, given what you just heard, what you know, and what you are thinking. Correspondingly, to learn a language is to learn to communicate with those who already speak it.

What gets communicated through the use of language is, of course, meaning — a tantalizingly intuitive concept that is easy to make precise (in a number of mathematically clever ways), but hard to make precise using formal tools that are (1) psychologically relevant, (2) neurobiologically plausible, and (3) most importantly, learnable from experience.[4] Not surprisingly, lowered expectations rule the day in semantics: "At least for now, the way to study meaning is by supposing that our publicly available sentences have meanings — and then trying to say how various features of sentences contribute to sentential meanings" (Pietroski, 2003).

Infants over the course of their development perceptibly progress from being, linguistically speaking, non-communicators to being experts at bending others to their will, in a gradual process whose rate and eventual degree of success depends critically on their sensorimotor activity and social environment (Goldstein et al., 2009). It makes sense, therefore, to ask how the "features of sentences" that contribute to their meanings can be learned from sensorimotor experience, **as a matter of principle**; in other words, what cues for learning to communicate are available in the raw data.[5]

To find that out, one must begin by subjecting the raw data — the utterances in some realistic corpus of experience, along with as many extralinguistic cues as are available — to a particular manipulation. In fact, what is called for here is precisely the same manipulation that constitutes the only possible basis for the discovery of any kind of structure in sequential data: the *alignment* of utterances to one another (that is, of the stream of data to shifted versions of itself) for the purposes of *comparison* (Harris, 1946, 1991; Solan et al., 2005; Edelman and Waterfall, 2007; Goldstein et al., 2009). Insofar as the raw data that are being subjected to this procedure are a record of embodied and physically and socially situated language use (and not just the "sound track" of the interaction), what a learner can glean from it are proper patterns of use — pragmatics and semantics, as it were, and not just syntax.

The data structure that best fits this notion of a record of experience and of how it should be processed is a kind of graph (Solan, Horn, Ruppin, and Edelman, 2005; Edelman, 2008a, p.274). Semi-formally, a graph is a discrete structure that consists of a set of vertices and a table that specifies which pairs of vertices are interconnected by edges. The set of discrete vertices in the present case may be found, for instance, in the phonemes of the language, whose sequence imposes a temporal order on all the rest of the information in a record of experience. Because the phonemes themselves can be extracted from raw speech data through alignment and comparison (Harris, 1946, 1952; see the review in Edelman, 2008a, ch.7), and because babies easily learn "words" formed by statistically stable patterns of phonemes (Saffran et al., 1996; Werker and Yeung, 2005), we may assume without loss of generality that the graph of experience is defined over words.[6]

The edges in this graph are directed: they are determined by the order of words in the utterances that comprise the corpus. The graph is heavily annotated by the various aspects of experience that label its vertices and edges: prosodic contours, pointers to the listener's conceptual structures, pointers to visual and other sensory information about the surrounding scene, social markers (including joint attention with and contingent feedback from the interlocutor(s)), records of motor acts, etc. (see (Goldstein et al., 2009) for a discussion of the importance of those cues in language acquisition).

This, then, is the fundamental structure of experience (minimally processed so as to impart to it a discrete sequential "backbone"), with which any cognitive agent (human, robotic, or alien) that sets out to learn to communicate with humans must contend. Such a graph structure can afford the system that harbors it only a minimal "look ahead" capability: the full range of documented continuations of a given utterance prefix is encoded in the graph, but the probability distribution over such continuations is still implicit. Moreover, the raw graph can support only limited comprehension (as in the mapping of a finite set of fully spelled-out utterances to conceptual or motor structures) and no productivity at all (no generation of novel utterances). In other words, merely committing experience to memory would allow the learner, at best, to act in some respects like a dog and in others like a parrot.

To go beyond one's own immediate experience and exhibit combinatorially open-ended comprehension and productivity, the listener must process and modify the graph. One way to do so is by recursively seeking partially alignable bundles of paths through it, thereby learning collocations, equivalences, and other statistical dependency patterns, which are assigned their own labeled vertices and are wired back into the graph. The result may be thought of as a kind of probabilistic "grammar" of sensorimotor experience, distilled from the original data. Solan et al. (2005) showed that such grammars learned automatically from raw transcripts of speech can be precise and productive — surprisingly so, given the highly impoverished nature of text-only data.

A much more concise and therefore powerful representation for a grammar of experience is a *higraph* — a directed graph in which subsets of the set of vertices may serve as vertices in their own right, edges may connect arbitrary tuples (rather than just pairs) of vertices, and Cartesian products of vertex sets are directly represented (Harel, 1988). A programming formalism for reactive systems based on higraphs, called *statecharts*, has proved to be widely applicable in computer science (Harel, 2007). It may help the reader to observe that a statechart bears the same relationship to a finite-state automaton as a higraph does to a graph, the former being exponentially more expressive; in other words, a finite-state machine whose behavior is equivalent to that of a given statechart may need exponentially more states (vertices).[7] A simple example of possible use of statecharts for integrated representation of speech and action — that is, syntax along with situated semantics — appears in Figure 1.

Computationally, the task of using sensorimotor experience to learn to communicate reduces, therefore, to the problem of distilling a statechart from a labeled graph that represents the raw data (a record of the learner's experience), subject to certain constraints, which must be specified as a part of the learning algorithm. Interestingly, statecharts have recently been taken up by game programmers, who use this formalism to specify patterns of discourse and behavior for computer game characters (Brusk, 2008). Within the present conceptual framework, this is an entirely expected turn. A game character's predicament resembles that of an infant in that it must make the best use of its limited experience and bounded computational resources to respond — preferably, on the basis of partial information, hence in an anticipatory manner — to the locutions and actions of the other characters, most importantly human players. Unfortunately, unsupervised algorithms for learning a statechart machine from samples of its intended behavior do not yet exist (except perhaps in babies' brains).[8] In developing such algorithms, every little helps. What we must consider next, then, is information that infants have at their disposal that helps them turn experience into practical, executable knowledge.

3 The richness of the stimulus

The problem of inferring a statechart (or any other kind of grammar) from samples of the behavior that it needs to generate is an instance of the wider problem of learning a (probabilistic) generative model for a set of (random-variable) data (Bishop, 2006; Hinton, 2007). Only very few unsupervised algorithms exist that are capable of working with raw transcribed language from large-scale realistic corpora, such as those in the CHILDES collection (MacWhinney, 2000); these are ADIOS (Solan et al., 2005), UDOP (Bod, 2009), and ConText (Waterfall et al., 2009). The performance of the grammars inferred by these algorithms cannot

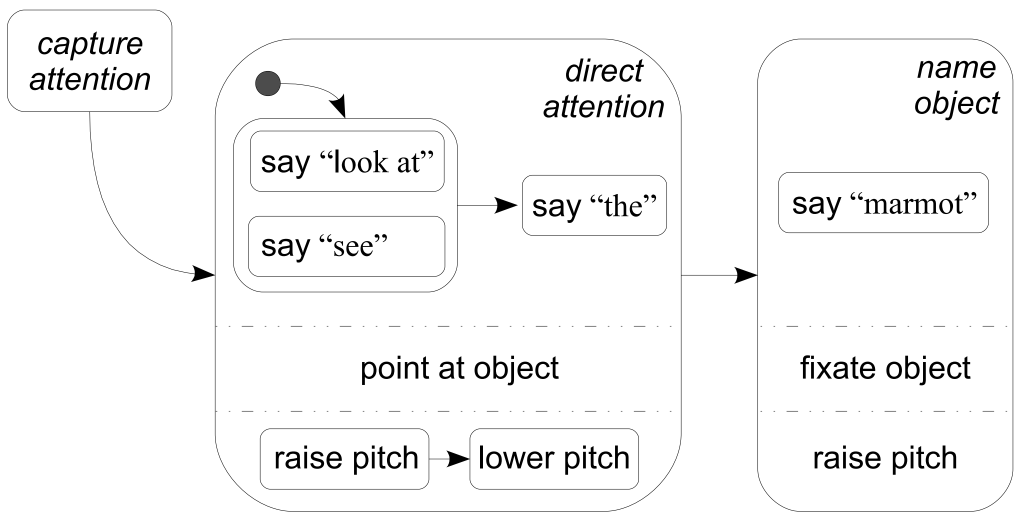


Figure 1: Statecharts are a powerful formalism for describing (or prescribing, if used generatively) behavior, which is based on the higraph notation (Harel, 1988). Informally, statecharts are state diagrams endowed with representational depth, orthogonality, and broadcast communication. The simple statechart in this example represents a routine for pointing out an object to a baby. It begins on the left with the capture of the baby's attention and proceeds simultaneously on three independent (orthogonal) tracks: lexical content (top), actions (middle), and prosody (bottom). The default entry point to the top of the *direct attention* node is at either of the two phrases, “look at” and “see,” which are mutually exclusive. In the *name object* node, the label “marmot” can be replaced with any member of the appropriate equivalence class (cf. Figure 2). For mathematical details, many illuminating examples, and pointers to literature where the syntax and the semantics of statecharts are rigorously defined, see (Harel, 1988).

yet compete with that of human learners. There is no doubt that this is due in part to the sparse sampling of language data that are available for learning (the conversations recorded in CHILDES are few and far apart, relatively to the density and total amount of speech to which a typical child is exposed). It would be instructive, however, to consider what characteristics of a child's language experience, apart from sheer volume, are not yet utilized by the state of the art learning algorithms.[9]

Quite tellingly, Smith and Gasser (2005), who offer "six lessons from babies" to those who seek to understand and perhaps emulate cognitive development, put language last: "starting as a baby grounded in a physical, social and linguistic world is crucial to the development of the flexible and inventive intelligence that characterizes humankind." In what follows, I briefly discuss three sources of cues, only one of which is linguistic, that likely assist development. These are the supra-sentential structure of discourse, the multimodal sensorimotor context that accompanies speech, and the dynamical social setting in which human linguistic interaction takes place.

3.1 Cross-sentential cues

In everyday child-directed speech, a large proportion of utterances come in the form of *variation sets* — runs of two or more sentences that share at least one lexical element (Küntay and Slobin, 1996; Waterfall, 2009; see Figure 2(a), top). A recent survey of the caregivers' parts of eight naturalistic interaction corpora from the English collection in CHILDES revealed this to be a pervasive phenomenon: over 20% of the utterances in the corpus occur within variation sets that contain at least two words in common. If a gap of up to two intervening sentences is allowed between two consecutive members of a variation set, this proportion rises to over 40% (when variation sets are defined by a single-word overlap, these figures rise to over 50% and 80%, respectively). Moreover, the lexical elements shared by the members of a typical variation set are not just some common function words: over 25% of unique words in the corpus participate in defining variation sets. These statistics apply to languages that are as different as Turkish, English, and Mandarin (Küntay and Slobin, 1996; Waterfall and Edelman, 2009).

Because of the partial lexical overlap, sentences in a variation set can be aligned, affording a natural way to compare them. Such comparison can yield informative and statistically reliable evidence of syntactic structure (Waterfall et al., 2009), and indeed longitudinal studies show that infants are better at structurally appropriate use of nouns and verbs that had occurred in their caregivers' speech within variation sets, compared to those that did not (Waterfall, 2006, 2009; cf. Nelson, 1977; Hoff-Ginsberg, 1986, 1990). An artificial grammar study with adult subjects confirmed the effectiveness of variation sets in making word segmentation and phrase structure easier to learn (Onnis et al., 2008).

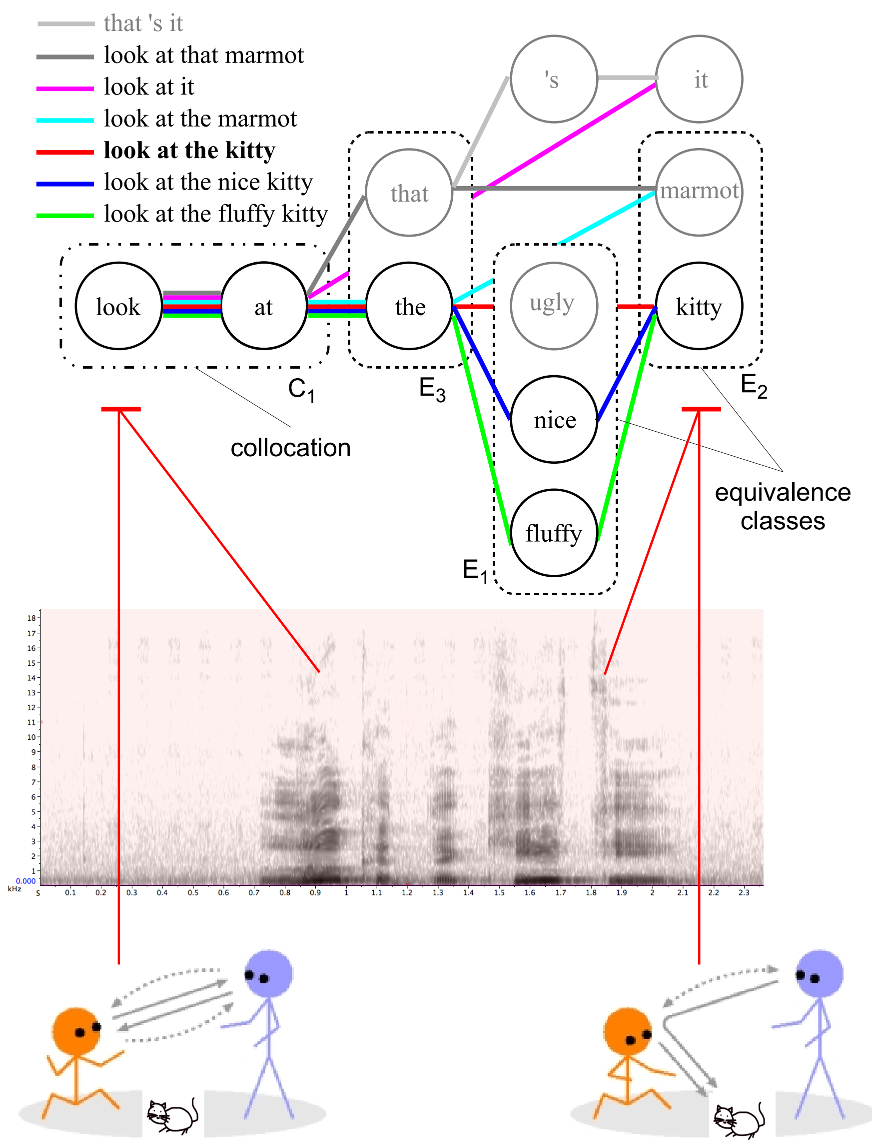
Most importantly for infant language acquisition, however, the manner in which variation sets bring out structure is local: the mechanisms of alignment and comparison need only span a few seconds' worth of the baby's attention, because the members of a variation sets are, by definition, never far apart from each other in time. Given how prevalent variation sets are, it is as if child-directed speech sets the baby up with an expectation of a partial repetition — and with it a morsel of certifiable structural knowledge about language — that is constantly renewed: each successive utterance is highly likely either to be continuing an already open variation set or to start a new one.

In its reliance on the anticipation of a partially familiar input, learning from variation sets takes advantage of predictive processing, a function which, as I pointed out in the opening section, language shares with other cognitive systems. Although variation sets become less prominent in caregivers' speech as the child grows older (Waterfall, 2006), partial overlap between utterances that occur in temporal proximity to each other in a conversation — that is, in naturally coordinated speech generated by two or more interlocutors — is extremely common (Du Bois, 2001, 2003; Szmrecsanyi, 2005). This realization, as well as abundant data on so-called syntactic priming (Bock, 1986; Bock et al., 2007), led researchers to speculate about the possible mechanisms that keep the participants in a conversation in tune with each other (Pickering and Garrod, 2004).

3.2 Multimodal cues

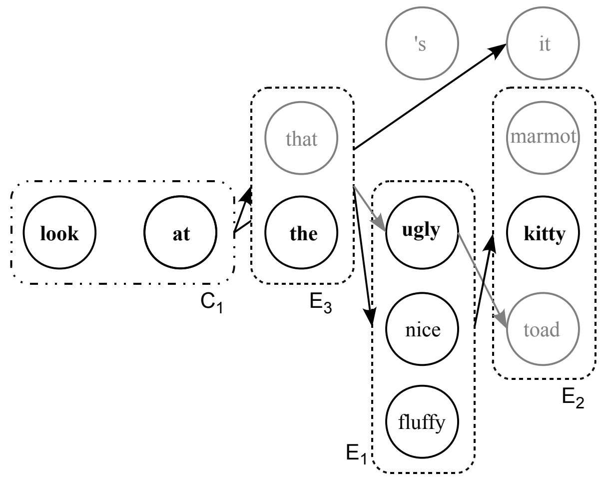
In addition to partial overlap between nearby chunks of speech, the statistical significance of patterns that emerge from data can be boosted by a convergence of multiple cues that join forces to highlight the same candidate structure. Such convergence is the first of the six principles of learning listed by Smith and Gasser (2005): "Babies' experience of the world is profoundly multi-modal. We propose that multiple overlapping and time-locked sensory systems enable the developing system to educate itself — without defined external tasks or teachers — just by perceiving and acting in the world." What William James (1890, p.488) described as an unruly mob of stimuli that beset the developing organism's senses ("the baby, assailed by eye, ear, nose, skin and entrails at once, feels it all as one great blooming, buzzing confusion") is in fact more like a well-organized circus parade in which troupes of mutually consistent cues reinforce each other and help the learner infer structure and impose order on its sensory experience (Goldstein et al., 2009).

The simplest example of such convergence in action can be found in word learning. As famously noted by Quine (1960), a mapping between a novel word and the object, if any, that it stands for cannot be inferred with complete confidence by mere observation. However, consistent cross-situational statistics do allow infants to learn a word-object pairing after encountering just a few "coincidences" the timing of each of



(a) From experience to grammar, PART I

Figure 2: *Top:* a small, 7-utterance, corpus and a graph that represents it, illustrating a **cross-sentential** (statistically significant alignment) cue to structure and meaning. *Middle:* a spectrogram of one of the utterances ("look at the kitty"), illustrating a **multimodal** (prosody + content) cue to structure and meaning. *Bottom:* eye fixation patterns, illustrating a **social** (joint attention) cue to structure and meaning. (Caption continued on the facing page.)



(b) From experience to grammar, PART II

Figure 2: (caption of Part I, continued from the facing page) — **Cross-sentential** cues. The seven utterances in this example, which is typical of child-directed speech, form a variation set (section 3.1). The way they overlap when aligned highlights certain statistical facts about this corpus, e.g., that "look at" is a collocation, C_1 and that "nice" and "fluffy" form an equivalence class. E_1, in the specific context shown (a separate bout of experience may indicate to the child that "ugly" belongs to E_1 too). **Multimodal** cues. The significance of these structures is boosted by the parallel stream of prosodic information (e.g., the rising pitch at "look" and at "kitty"). **Social** cues. Joint attention combines with alignment and prosody to single out the label "kitty" and to fix its situational reference (i.e., the cat, which the baby and the caregiver both end up looking at; the drawings are courtesy of Hideki Kozima, Miyagi University, Japan). PART II — The graph in (a), top, can be distilled into this compressed form (conceptual and other extralinguistic annotations, without which the grammar would be worthless as a guide for behavior, have been omitted for clarity). Observe that one arrow connects E_3 to E_1 (that is, to any of its members) and another connects it just to "ugly" (which thereby is distinguished from other members of E_3). This representation affords productivity: it can generate the utterance "look at the ugly marmot" that the learner may never have encountered. A formalism that includes this and several others, much more powerful, representational tools is statecharts (see section 2). The statechart notation may seem overly complicated compared to the standard one (e.g., $S \rightarrow C_1 E_3 E_1 E_2$; this slightly too lax rule would ensue if the distinction between the contextual expectations of "ugly" on the one hand and "nice" and "fluffy" on the other hand were dropped). However, speech is situated not on a book page but in the world (which has no place in the standard notation) and it starts not with the empty symbol S but rather with an activation of some concepts, such as look and kitty, which then spreads through the graph until an utterance is ready for output. The statechart formalism fits these requirements to the T.

which if taken in isolation is insufficiently precise (Smith and Yu, 2008). The power of such statistical learning is further increased when additional cues, such as the speaker's prosody and joint attention between the speaker and the listener, are utilized (Figure 2(a), middle and bottom; cf. Yu and Ballard, 2007). The use of multiple cues does not, of course, cease in adulthood: there is now massive evidence to the effect that language processing during comprehension is incremental and relies heavily on a plethora of perceptual cues assisting each other's interpretation through cross-modal expectations (Crocker et al., 2009).

3.3 Social cues

Being a key component of the human "interactional engine" (Levinson, 2006), language is closely intertwined with other social communication mechanisms that are available to people. It obviously facilitates social cognition, and just about every other kind of cognition as well, thereby serving as a scaffolding for the growth of the human civilization (Clark, 1998). It is becoming increasingly clear that this facilitation is bidirectional. As argued by Herrmann et al. (2007), the distinctly human social traits, many of which are exhibited already by prelinguistic babies, may have been essential for the emergence of human language, whose evolution proceeds at a much faster pace than the evolution of its host species (Christiansen and Chater, 2008). It stands to reason, therefore, that social factors should shape language development, and indeed they do (Hoff, 2006).

The social origin of many of the multimodal cues mentioned earlier, such as the characteristic prosody of child-directed speech (Yu and Ballard, 2007; Pereira et al., 2008), is but one aspect of this influence. A much more powerful mechanism through which caregivers can shape and facilitate learning is social feedback that is contingent on the baby's own actions. As my colleagues and I have argued elsewhere (Goldstein et al., 2009), such social interaction allows candidate linguistic structures to stand out from a continuous stream of experience by passing two kinds of significance tests.

The first of these tests is intrinsic to the speech data; it applies, for example, when partially alignable chunks of utterances in a variation set highlight a structure that may be worth learning (as noted earlier in this section). The second test is socially situated: by definition for a communication system, "interesting" structures must be behaviorally significant, as indicated by cues that are extrinsic to the stream of speech. There is growing evidence that socially guided learning that relies on both tests provides a powerful early impetus to the language acquisition process (Goldstein et al., 2003; Goldstein and Schwade, 2008).[10] In this connection, we may observe that social feedback works by facilitating the delivery of information precisely when the baby *expects* it (and is therefore self-motivated to give it due processing diligence).

4 Sensorimotor experience and the brain

In this chapter, I have already deviated twice from the usual practice of formalist linguistics of approaching the study of language with a firm preconception of what the answers to the big questions should look like. First, I identified an open inquiry into the informational structure of experience, of which speech is but one strand, as a prerequisite for any study of "grammar," conceived of properly as a distillation of experience (Figure 2(b)). Second, by focusing on the information that is made available to infants by their caregivers and environment, I noted three clusters of properties that can facilitate the distillation of experience into a vehicle of anticipation (in listening) and eventual purposeful production (in speaking).

For good measure, I shall now commit a third transgression: instead of acquiescing to the textbook assertion that there exists in the brain a language module whose evolutionary origins, developmental trajectory, and neurocomputational circuitry can all be left safely and conveniently opaque,[11] I shall line up and discuss, necessarily briefly, a series of insights and findings from brain science that language theorists can ill afford to ignore. The main thrust of the discussion will be to argue that language acquisition and use involve certain general-purpose (i.e., not exclusively linguistic) functions of the "language" areas in the frontal lobe of the cortex and, more importantly, of certain subcortical structures (Lieberman, 2002; Müller and Basho, 2004; Ullman, 2006).

4.1 The hippocampus

Let us first consider the hippocampus, a subcortical structure that resides in the medial temporal lobe (MTL) of the brain. Classical studies in rats, beginning with O'Keefe and Dostrovsky (1971), led to the common view of the hippocampus as a cognitive map. Subsequent research showed that its function is predictive (Muller and Kubie, 1989) and that it is map-like in that it integrates a wide range of episodic information about spatially anchored events (Eichenbaum et al., 1999). More recently, it became apparent that memory traces for *sequences* of events are laid down in the hippocampus and that both the events and their ordering may be abstract rather than spatial (Levy, 1996; Fortin et al., 2002; Levy et al., 2005; Buzsáki, 2005). The role of the hippocampus in learning sequence-structured data is especially important (i) when the sequences partially overlap, so that each distinct prefix of a common subsequence determines its respective distinct suffix (Levy, 1996) and (ii) a substantial amount of time may elapse between successive items in a sequence (Agster et al., 2002).

The view of the hippocampus that emerges from the animal studies is that of a computational tool that is honed to process multimodal (sensorimotor plus abstract) graph-structured data, which would make it well-suited to handle the distillation of experience into a probabilistic statechart grammar. This notion does not really contradict the established view of the role of the hippocampal formation in humans, which holds it to be essential for explicit memory and for long-term memory consolidation: what is language learning if not a massive exercise in memory consolidation? Once we begin seeing the hippocampus in this light, several pieces of the puzzle fall into place.

First, the hippocampus is presumably a key brain structure that makes space matter in discourse. Just as rats learn better when the data are presented to them in a spatially consistent manner, human infants are better at word learning when location is used consistently to anchor word reference (Hockema and Smith, 2009). This developmental finding complements the results of behavioral studies with adults that show similarly strong effects of space serving as a scaffolding for building up bundles of episodic information (Richardson and Spivey, 2000) and as a medium for dynamic social coordination between interlocutors (Richardson and Dale, 2005).

Second, the hippocampus is involved in language processing. This is suggested by its role in implicit sequence learning, of the kind that psychologists test with small artificial grammars. The expectation that such tests should be relevant to the processing of real language is borne out both by EEG and by fMRI imaging results (Meyer et al., 2005; Schendan et al., 2003). Furthermore, imaging studies show that hippocampal activity distinguishes between good and poor learners of sequence tasks (Breitenstein et al., 2005). Individual variability in implicit sequence learning also correlates with performance in the processing of sequential context in spoken language (Conway and Pisoni, 2008).

Third, the hippocampus appears to be indispensable for language acquisition. Thus, partial lesions of the hippocampus result in developmental amnesia, in which the patient exhibits in particular a reduced ability to recall sequentially structured information after a 24-hour delay (Adlam et al., 2005). Early left hippocampal pathology results in abnormal language lateralization (Weber et al., 2006). Most tellingly, infants who suffer from extensive bilateral hippocampal sclerosis early in life fail to acquire language (or lose attained language) or to develop social and adaptive skills, despite adequate sensorimotor functioning (DeLong and Heinz, 1997).

4.2 The basal ganglia

Extraordinary feats of memory require extraordinary motivation, as well as proper coordination between data compression,[12] sequencing, and associative storage mechanisms. One would expect that the social cues that highlight complex sequential structure in the stream of experience would also help motivate learning, and that mechanisms of motivated sequence learning could be shared between all the tasks that need them (Lashley, 1951). This is indeed the case; in addition to areas in the medial temporal lobe (the hippocampus and the entorhinal cortex), in the frontal lobe, and in the thalamus, the system in question includes, most significantly, the basal ganglia.

Behavioral and neuropsychological findings in humans show that the basal ganglia interact with the hippocampus and with cortical areas in supporting learning and execution of a wide variety of cognitive tasks that require flexible coordination of sequential structure processing and working memory (Seger, 2006), including language (Lieberman, 2002, pp.116-119). Ullman (2006, p.482) suggests that "the basal ganglia may play a particularly important role in the acquisition of grammatical and other procedural knowledge, whose use eventually depends largely on the posterior portion of Broca's area." Moreover, the basal ganglia circuits also handle the social-motivational aspects of complex learning, in all species that are capable of it (Syal and Finlay, 2009). Although this system receives much attention from neuroscientists and from computational modelers (Dominey, 2005; Dominey and Hoen, 2006; O'Reilly and Frank, 2006; Cohen and Frank, 2009),[13], the social computing role of basal ganglia is rarely mentioned. Given how important this system is, one hopes that before long "researchers in early language development turn their attention from the storage device, the cortex, to the neuroanatomy which provide[s] the motivational structure for behavior, the basal forebrain and striatum" (Syal and Finlay, 2009).

4.3 All together now

In mammals, the hippocampus sits at the functional apex of three converging bidirectional streams of information, which are channeled by somatosensory-motor, visual, and auditory isocortical hierarchies, respectively (Merker, 2004). Furthermore, the hippocampus and the basal ganglia have bidirectional functional links to the prefrontal cortex (Okanoya and Merker, 2007, fig. 22.4), an arrangement that is starting to attract modeling efforts (O'Reilly and Norman, 2002; O'Reilly, 2006).[14] Coordination among all these brain structures is required for learning sequential behaviors, for exerting control over their production, and for committing them to long-term memory (Shapiro, 2009).

The mechanisms of this coordination are being thoroughly studied in animals. For instance, there is much evidence for replay of maze traversal experience during sleep in rats (Lee and Wilson, 2002), which is analogous to song replay during sleep in songbirds (Dave and Margoliash, 2000). Such replay, whose unfolding is coordinated between the hippocampus and the cortex (visual and prefrontal; Ji and Wilson, 2007; Peyrache, Khamassi, Benchenane, Wiener, and Battaglia, 2009) is thought to underlie memory consolidation. The coordination is mediated by oscillatory activity (Jones and Wilson, 2005), whose frequency and phase relationships across regions are tightly controlled (Buzsáki, 2010). Imaging evidence is becoming available that favors the existence in the human brain of an analogous system for sequence learning, consolidation, and production (Schendan et al., 2003; Seger and Cincotta, 2006; Albouy et al., 2008).[15]

5 Language acquisition reloaded

The mosaic of behavioral, electrophysiological, and imaging findings surveyed in sections 3 and 4 is consistent with the theoretical framework that I outlined earlier that addresses the initial representation of experience and its eventual distillation into a form of generative grammar that in humans supports all complex, hierarchically structured, sequential behaviors, including language. Much work remains to be done, both in integrating the wealth of experimental findings and in developing a viable neurocomputational approach to statechart learning that would draw on the integrated empirical data.[16] The open issues that remain cannot even be all listed, let alone resolved, here, which is why I shall offer merely a sample of one question each on the problem, algorithm, and mechanism levels.

5.1 The big statechart in the sky

The first question pertains to the scope of the statechart grammar that a situated language user is expected to require. As implied by a theorem proved by Conant and Ashby (1970), a cognitive system that aims to control its fate must maintain an internal model of its environment. This model, as noted earlier, must be probabilistic and generative, to better deal with the ubiquitous and unavoidable uncertainties. In a society of cognitive agents, of which a linguistic community is a special case, an individual's internal model must, therefore, include both the shared environment and other agents.[17]

The methodological virtues of this approach have been discussed by Syal and Finlay (2009), who conclude: "In the avian species that display learned vocal behavior, the learning environment is an integrated system, viewed best when the entire infant bird, its tutor, the interaction between the two, and the effect of each actor on its own, and the other's nervous system, are considered." In humans, arguments for socially shared representations have been put forward by Decety and Sommerville (2003); in the case of language, it is hypothesized that such representations involve emulating the other speaker in a dialogue (Pickering and Garrod, 2007) (for a survey of the available evidence and computational arguments, see Edelman, 2008a, ch.6,7). The statechart formalism, which has been developed to model reactive systems and which has powerful means for representing combinatorially complex, nested relationships, is well-suited for capturing grammars that involve multiple interacting agents.

5.2 The ace of Bayes?

The second question is how to learn such grammars from records of experience. I believe that the answer to this question will be Bayesian. A Bayesian foundation for word learning has been described by Frank et al. (2009), who showed that considering simultaneously word reference fixation and the speaker's referential intention is more effective than treating each problem separately. More generally, Bayesian cue integration is also the proper approach to multimodal perception (Kersten and Yuille, 2003) and, indeed, to any kind of learning and inference from data in cognition (Chater et al., 2006; Edelman, 2008a).

In the ADIOS algorithm for distilling a grammar from a graph representing a corpus of language (Solan et al., 2005), the criterion for rewiring the graph relied on a simple binomial significance test for vertex connectivity. Clearly, we must do better than that. A hint as to how a Bayesian approach could be made to work for this problem can be found in recent computational analysis of experience-based modification of the hippocampal network by Levy et al. (2005, p.1252), who noted that the use of Bayesian "inversion" allows the active graph formed by the CA3-entorhinal loop to estimate forward-looking dependencies — that is, formulate predictions — as a part of its processing of past experiences.

A very general abstract Bayesian model for learning graphs (or any other structural representations) from relational data has been recently proposed by Kemp and Tenenbaum (2008). The worked examples they offer begin with writing down a multivariate Gaussian with a dimension for each node in the graph to parametrize the generative model, and proceed by performing a greedy search, guided by the Bayes formula, in the resulting parameter space. Although in principle this approach can be applied as is to the problem of statechart learning, scaling is bound to become a problem with realistic corpora of experience. An intriguing possibility for resolving the scaling issue is to try to isolate "islands" in the statechart grammar where learning can be made effectively local, perhaps under the influence of variation sets and other local cues in the data.

5.3 Time and again

In the cortex, which is where the distilled grammar would be anchored in long-term memory according to the present framework (cf. Ullman, 2006, p.482), the dynamic representation of a particular sequence of states (say, the phonemes that form a word) may take the form of a synfire chain — an orderly propagation of activity between designated cliques of neurons (Abeles, 1982; Bienenstock, 1992). Evidence for the existence of such chains of activity in the primary visual cortex (Ikegaya et al., 2004) indicates that synfire-based representations are biologically feasible. Indeed, synfire activity arises spontaneously (Izhikevich, 2006), or in response to input patterns (Hosaka et al., 2008), in recurrent networks of simulated spiking neurons that learn from the statistics of their own firing experience via a spike timing-dependent plasticity (STDP) synaptic modification rule.[18]

A pressing implementational issue that needs to be resolved for STDP to provide an explanation of the synaptic mechanism of learning language is that of timing (cf. Wallenstein, Eichenbaum, and Hasselmo, 1998, p.318). STDP operates on the time scale of tens of milliseconds at most; in comparison, language (and human behavior in general) unfolds on the time scale of seconds, while social and other cues that are contingent on one's utterance or act may not come until much later. Izhikevich (2007) showed that this issue can be addressed through combined action of a fast STDP mechanism that "marks" the relevant synapses and a subsequent slow process of stabilizing the change, which depends on reward-related release of dopamine — a neurotransmitter that mediates learning in the loops connecting the basal ganglia with the cortex (Da Cunha et al., 2009).

A similar issue arises in understanding memory replay in the rat (Ji and Wilson, 2007): the replay, which presumably happens to allow STDP-based consolidation, is much faster than the animal's running speed, indicating that some form of time compression takes place (Jensen and Lisman, 2005; Davidson et al., 2009). Interestingly, the manner in which time compression in memory replay works may hold a hint as to how the hippocampus-cortex-striatum network can learn and generate long sequences that are formed by various combinations of shorter ones. Replay in the rat hippocampus coincides with high-frequency "ripple" oscillations, which do not last long enough to represent long treks through the rat's environment, but, as shown by Davidson et al. (2009), are composed combinatorially, thus altering the behavioral meaning of the entire event.[19] Moreover, Davidson et al. (2009, p.504) noted that "replayed trajectories represent the set of possible future or past paths linked to the animal's current position rather than the actual paths." Add to this capability a modicum of recursion (by allowing some subsequences to be nested within others, up to a point), and you have a biological substrate for complex behavior, including language.

6 Conclusion

I have now come a full circle to the beginning of this chapter — back to the twin notions that, first, mastering language must have something to do with getting *meaning* in and out of it and, second, that meaning must have something to do with the way language is *used* as part of behavior. In doing so, I could not help noticing some intriguing parallels between a computational analysis of the nature of linguistic experience and the deep insights into this very same matter that are to be found in the work of Ludwig Wittgenstein — a philosopher whose construal of meaning is usually condensed to the maxim "meaning is use," which has been misused to the point of meaninglessness.

A better idea of what Wittgenstein may have had in mind can be obtained by considering three extraordinary passages from *Philosophical Investigations*. The first one, which I used as the motto for this chapter, broaches the possibility that a record of linguistic experience may look like a graph ("A multitude of familiar paths lead off from these words in every direction"). The second one reiterates this view of language and connects it to both vision and action:

> Phrased *like this*, emphasized like this, heard in this way, this sentence is the first of a series in which a transition is made to *these* sentences, pictures, actions. ((A multitude of familiar paths lead off from these words in every direction.))
>
> — Wittgenstein (1958, §534, p.142)

Finally, in the third passage Wittgenstein rounds off his observation with some memorable metaphors:

> Suppose someone said: every familiar word, in a book for example, actually carries an atmosphere with it in our minds, a 'corona' of lightly indicated uses. — Just as if each figure in a painting were surrounded by delicate shadowy drawings of scenes, as it were in another dimension, and in them we saw the figures in different contexts.
>
> — Wittgenstein (1958, II:VI, p.181)

What a language user relies upon in looking ahead to the successive installments of the incoming utterance, or in constructing a sequence of words to be uttered, is a probabilistically annotated graph-like record of experience — the "multitude of familiar paths" along with the "'corona' of lightly indicated uses" — which has been incorporated into the general "grammar" that drives behavior. Developing biologically relevant algorithms that can distill multimodal experience in this manner is the great challenge that the research program I sketched here will have to tackle next.

Acknowledgments

The idea of representing a corpus of language in the form of a graph arose in a series of conversations that I had with Zach Solan in 2001-2003. The final form of section 4 was influenced by discussions I participated in at the Ernst Strüngmann Forum on Dynamic Coordination in the Brain, held in August 2009 at the Frankfurt Institute for Advanced Studies. I thank Barb Finlay, Mike Goldstein, David Harel, David Horn, Björn Merker, Luca Onnis, Geoff Pullum, Eytan Ruppin, Ben Sandbank, Jen Schwade, Aaron Sloman, Mark Steedman, Heidi Waterfall, and Matt Wilson for comments on various aspects of this project.

Notes

[1]In this connection, we may consider the debunking by Pereira (2000) of Chomsky's claim of irrelevance of statistics to language that is based on his famous "colorless green ideas sleep furiously" example: a simple corpus-based statistical model of language handily labeled this sentence as $200,000$ times more probable than its scrambled version.

[2]Although the "generative" label has been traditionally associated exclusively with the Chomskian brand of linguistics, in reality it applies to any approach that calls for learning a generative probabilistic model of a data set — an empiricist notion *par excellence* (Goldsmith, 2007) and the only universally valid way of dealing with data that affords generalization (Bishop, 2006). For an outline of an empirical generative framework for understanding language acquisition, see (Waterfall et al., 2009).

[3]An early expression of the conviction that "syntax" is an independent level that, moreover, cannot be sidestepped is offered by Chomsky (1957, p.87): "What we are suggesting is that the notion of 'understanding a sentence' be explained in part in terms of the notion of 'linguistic level.' To understand a sentence, then, it is first necessary to reconstruct its analysis on each linguistic level." The supposed "autonomy of syntax" has been recently reaffirmed by Chomsky (2004, p.138).

[4]Some progress has been made in modeling human processing of meaning in various circumscribed situations, such as dealing with simple logical problems (Stenning and van Lambalgen, 2008). In computational linguistics, the learning of semantics is either heavily supervised (e.g., the wide-coverage semantic parser of Bos et al. (2004) works from very detailed semantic knowledge that's built into its lexicon-grammar) or else works for highly simplified situations (e.g., Eisenstein, Clarke, Goldwasser, and Roth, 2009).

[5]This formulation of the question stresses that it pertains to what Marr and Poggio (1977) termed the abstract computational level. Note that the popular trick of declaring it all innate amounts to dodging the question rather than answering it (Putnam, 1967).

[6]At the level that matters, language is "digital" (that is, defined over a set of discrete primitives) for reasons of computational tractability (Edelman, 2008b).

[7]For once, Chomsky (2004, p.92) gets it right: "It is obvious, in some sense, that processing systems are going to be represented by finite state transducers. That has got to be the case [. . .] But that leaves quite open the question of what is the internal organization of the system of knowledge."

[8]In one of the existing algorithms, a teacher (the designer) serves as an oracle that evaluates pieces of generated behavior and decides the fate of the rules that gave rise to them (Mäkinen and Systä, 2002). In another work, statecharts are synthesized from scenario-based requirements, themselves stated in a formal language (Harel et al., 2005).

[9]Cf. Bates et al. (1999): "Consider the following statistics: assuming a taciturn Calvinist family in which an English-speaking child hears approximately 5 hours of speech input per day, at a mean rate of 225 words per minute, the average 10-year-old child has heard $1,034,775,000$ English phonemes (at an average of $25,869,375$ trials per phoneme). She has heard just under 250 million words (including $17,246,250$ renditions of the most common function words) and 28 million sentences [. . .]."

[10]Social guidance also helps robots learn to solve puzzles (Thomaz and Breazeal, 2008).

[11]Cf. Chomsky (2004, p.56): "I think a linguist can do a perfectly good work in generative grammar without ever caring about questions of physical realism or what his work has to do with the structure of the mind."

[12]Data compression is critically important not only because of capacity limitations: without compression, there can be no generalization and therefore no prediction ability (Grünwald, 1994).

[13]An entire special issue of of *Behavioural Brain Research* (volume 199, number 1, 2009) was devoted to the role of basal ganglia in learning and memory.

[14]The isocortical and especially the frontal areas, are, of course, much more extensive in humans than in other mammals, which

explains, in part, why not all species that have the "standard" machinery in place for processing sequences (thalamus, basal ganglia, hippocampus, prefrontal cortex) can learn to play the violin or engage in conversation. Merker and Okanoya (2007) relate the emergence of language to encephalization in humans.

[15]In what must count as an understatement of the year, Albouy et al. (2008) write: "Motor sequences constitute an integral part of a number of everyday life activities such as writing, typing, speaking, knitting, or playing a musical instrument." !

[16]The idea of a "grammar of behavior" is related to the notion of action schemata, which has been entertained by psychologists for some decades now (Lashley, 1951; Arbib, 2006). Houghton and Hartley (1996) offer a particularly cogent discussion of the obstacles that any neurocomputational implementation of serially and hierarchically structured schemata must overcome.

[17]In modeling other agents, one must beware of too deep a recursion, as illustrated by the following excerpt from the script of *The Princess Bride* by William Goldman:

> MAN IN BLACK
> All right: where is the poison? The battle of wits has begun. It ends when you decide and we both drink, and find out who is right and who is dead.
> VIZZINI
> But it's so simple. All I have to do is divine from what I know of you. Are you the sort of man who would put the poison into his own goblet, or his enemy's? Now, a clever man would put the poison into his own goblet, because he would know that only a great fool would reach for what he was given. I'm not a great fool, so I can clearly not choose the wine in front of you. But you must have known I was not a great fool; you would have counted on it, so I can clearly not choose the wine in front of me.

(This goes on for a bit longer before one of them dies.)

[18]STDP is a Hebbian learning rule (Caporale and Dan, 2008), which has interesting connections to Bayesian inference (Deneve, 2008).

[19]Henson and Burgess (1997) hypothesized that sequential information could be represented in the brain by a collection of oscillators operating at different frequencies. Specifically, they showed that a sequence can be coded by the oscillator whose half-period best fits its length, with the position of each item in the sequence being signaled by the phase of the oscillator at the point in time when that item was presented. In addition to accounting for a range of behavioral data on memory for sequences (Henson, 1999), this model fits well the recent findings on the interplay of various oscillatory regimes in the hippocampus (Davidson et al., 2009).

References

Abeles, M. (1982). Role of cortical neuron: integrator or coincidence detector? *Israel J. Med. Sci. 18*, 83–92.

Adlam, A.-L. R., F. Vargha-Khadem, M. Mishkin, and M. de Haan (2005). Deferred imitation of action sequences in developmental amnesia. *Journal of Cognitive Neuroscience 17*, 240–248.

Agster, K. L., N. J. Fortin, and H. Eichenbaum (2002). The hippocampus and disambiguation of overlapping sequences. *The Journal of Neuroscience 22*, 5760–5768.

Albouy, G., V. Sterpenich, E. Balteau, G. Vandewalle, M. Desseilles, T. Dang-Vu, A. Darsaud, P. Ruby, P.-H. Luppi, C. Degueldre, P. Peigneux, A. Luxen, and P. Maquet (2008). Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron 58*, 261–272.

Arbib, M. A. (2006). A sentence is to speech as what is to action? *Cortex 42*, 507–514.

Baker, J. K. (1975). Stochastic modeling for automatic speech understanding. In R. Reddy (Ed.), *Speech Recognition*, pp. 521–542. Academic Press.

Bates, E. and J. C. Goodman (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *Emergence of Language*, pp. 29–79. Hillsdale, NJ: Lawrence Earlbaum Associates.

Bates, E., D. Thal, B. L. Finlay, and B. Clancy (1999). Early language development and its neural correlates. In I. Rapin and S. Segalowitz (Eds.), *Handbook of neuropsychology* (2 ed.), Volume 7. Amsterdam: Elsevier.

Bienenstock, E. (1992). Suggestions for a neurobiological approach to syntax. In D. Andler, E. Bienenstock, and B. Laks (Eds.), *Proceedings of Second Interdisciplinary Workshop on Compositionality in Cognition and Neural Networks*, pp. 13–21. Abbaye de Royaumont, France.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Berlin: Springer.

Bloom, L. (1970). *Language development: form and function in emerging grammars*. Cambridge, MA: MIT Press.

Bock, J. K. (1986). Syntactic priming in language production. *Cognitive Psychology 18*, 355–387.

Bock, K., G. S. Dell, F. Chang, and K. H. Onishi (2007). Persistent structural priming from language comprehension to language production. *Cognition 104*, 437–458.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science 33*, 752–793.

Bos, J., S. Clark, J. R. Curran, J. Hockenmaier, and M. Steedman (2004). Wide-coverage semantic representations from a CCG parser. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, Geneva, Switzerland.

Breitenstein, C., A. Jansen, M. Deppe, A.-F. Foerster, J. Sommer, T. Wolbers, and S. Knecht (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage 25*, 958–968.

Brusk, J. (2008). Dialogue management for social game characters using statecharts. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*, Yokohama, Japan, pp. 219–222.

Buzsáki, G. (2005). Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus 15*, 827–840.

Buzsáki, G. (2010). Oscillation-supported information processing and transfer in the hippocampus-entorhinal-neocortical interface. In C. von der Malsburg, W. A. Phillips, and W. Singer (Eds.), *Dynamic Coordination in the Brain: From Neurons to Mind*, Volume 5 of *Strüngmann Forum Report*. Cambridge, MA: MIT Press.

Caporale, N. and Y. Dan (2008). Spike timing-dependent plasticity: A Hebbian learning rule. *Annual Review of Neuroscience 31*, 25–46.

Chater, N., J. B. Tenenbaum, and A. Yuille (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences 10*, 287–291.

Chipere, N. (2001). Native speaker variations in syntactic competence: implications for first language teaching. *Language Awareness 10*, 107–124.

Chomsky, N. (1957). *Syntactic Structures*. the Hague: Mouton.

Chomsky, N. (2004). *The Generative Enterprise Revisited*. Berlin: Mouton de Gruyter. Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi.

Christiansen, M. H. and N. Chater (2008). Language as shaped by the brain. *Behavioral and Brain Sciences 31*, 489–509.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers and J. Boucher (Eds.), *Language and thought: Interdisciplinary themes*, pp. 162–183. Cambridge: Cambridge University Press.

Cohen, M. X. and M. J. Frank (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. *Behavioural Brain Research 199*, 141–156.

Conant, R. C. and W. R. Ashby (1970). Every good regulator of a system must be a model of that system. *Intl. J. Systems Science 1*, 89–97.

Conway, C. M. and D. B. Pisoni (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Science 1145*, 113–131.

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge, England: Cambridge University Press.

Crocker, M. W., P. Knoeferle, and M. R. Mayberry (2009). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*. In press.

Da Cunha, C., E. C. Wietzikoski, P. Dombrowski, M. Bortolanza, L. M. Santos, S. L. Boschen, and E. Miyoshi (2009). Learning processing in the basal ganglia: A mosaic of broken mirrors. *Behavioural Brain Research 199*, 157–170.

Dabrowska, E. and J. Street (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences 28*, 604–615.

Dave, A. S. and D. Margoliash (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science 290*, 812–816.

Davidson, T. J., F. Kloosterman, and M. A. Wilson (2009). Hippocampal replay of extended experience. *Neuron 63*, 497–507.

Decety, J. and J. A. Sommerville (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences 7*, 527–533.

DeLong, G. R. and E. R. Heinz (1997). The clinical syndrome of early-life bilateral hippocampal sclerosis. *Annals of Neurology 42*, 11–17.

Deneve, S. (2008). Bayesian spiking neurons I: Inference. *Neural Computation 20*, 91–117.

Dennett, D. C. (2003). *Freedom evolves*. New York: Viking.

Dewey, J. (1910). *How we think*. Lexington, MA: D. C. Heath.

Diessel, H. (2004). *The Acquisition of Complex Sentences*, Volume 105 of *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.

Dominey, P. F. (2005). From sensorimotor sequence to grammatical construction: Evidence from simulation and neurophysiology. *Adaptive Behavior 13*, 347–361.

Dominey, P. F. and M. Hoen (2006). Structure mapping and semantic integration in a construction-based neurolinguistic model of sentence processing. *Cortex 42*, 476–479.

Du Bois, J. W. (2001). Towards a dialogic syntax. Unpublished manuscript.

Du Bois, J. W. (2003). Argument structure: Grammar in use. In J. W. Du Bois, L. E. Kumpf, and W. J. Ashby (Eds.), *Preferred Argument Structure. Grammar as Architecture for Function*, pp. 11–60. Amsterdam: John Benjamins.

Edelman, S. (2008a). *Computing the mind: how the mind really works*. New York: Oxford University Press.

Edelman, S. (2008b). On the nature of minds, or: Truth and consequences. *Journal of Experimental and Theoretical AI 20*, 181–196.

Edelman, S. and H. R. Waterfall (2007). Behavioral and computational aspects of language and its acquisition. *Physics of Life Reviews 4*, 253–277.

Eichenbaum, H., P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila (1999). The hippocampus, memory and place cells: is it spatial memory or memory space? *Neuron 23*, 209–226.

Eisenstein, J., J. Clarke, D. Goldwasser, and D. Roth (2009). Reading to learn: Constructing features from semantic abstracts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, Singapore.

Fortin, N. J., K. L. Agster, and H. B. Eichenbaum (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience 5*, 458–462.

Frank, M. C., N. D. Goodman, and J. B. Tenenbaum (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science 20*, 578–585.

Goldsmith, J. A. (2007). Towards a new empiricism. In J. B. de Carvalho (Ed.), *Recherches linguistiques à Vincennes*, Volume 36.

Goldstein, M. H., A. P. King, and M. J. West (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Science 100*, 8030–8035.

Goldstein, M. H. and J. A. Schwade (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science 19*, 515–523.

Goldstein, M. H., H. R. Waterfall, A. Lotem, J. Halpern, J. Schwade, L. Onnis, and S. Edelman (2009). General cognitive principles for learning structure in time and space. Submitted.

Goodman, J. T. (2001). A bit of progress in language modeling. *Computer Speech and Language 15*, 403–434.

Grünwald, P. (1994). A minimum description length approach to grammar inference. In G. Scheler, S. Wernter, and E. Riloff (Eds.), *Connectionist, statistical and symbolic approaches to learning for natural language*, Volume 1004 of *Lecture Notes in AI*, pp. 203–216. Berlin: Springer Verlag.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science 30*, 643–672.

Harel, D. (1988). On visual formalisms. *Commun. ACM 31*, 514–530.

Harel, D. (2007). Statecharts in the making: a personal account. In *HOPL III: Proceedings of the third ACM SIGPLAN conference on History of programming languages*, New York, NY, pp. 5–1–5–43. ACM.

Harel, D., H. Kugler, and A. Pnueli (2005). Synthesis revisited: Generating statechart models from scenario-based requirements. In *Lecture Notes in Computer Science*, Volume 3393, pp. 309–324. Springer-Verlag.

Harris, Z. S. (1946). From morpheme to utterance. *Language 22*, 161–183.

Harris, Z. S. (1952). Discourse analysis. *Language 28*, 1–30.

Harris, Z. S. (1991). *A theory of language and information*. Oxford: Clarendon Press.

Henson, R. N. A. (1999). Coding position in short-term memory. *International Journal of Psychology 34*, 403–409.

Henson, R. N. A. and N. Burgess (1997). Representations of serial order. In J. A. Bullinaria, D. W. Glasspool, and G. Houghton (Eds.), *4th Neural Computation and Psychology Workshop*, pp. 283–300. London: Springer.

Herrmann, E., J. Call, M. Lloreda, B. Hare, and M. Tomasello (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science 317*, 1360–1366.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences 11*, 428–434.

Hockema, S. A. and L. B. Smith (2009). Learning your language, outside-in and inside-out. *Linguistics 47*, 453–479.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review 26*, 55–88.

Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: their relation to the child's development of syntax. *Developmental Psychology 22*, 155–163.

Hoff-Ginsberg, E. (1990). Maternal speech and the child's development of syntax: A further look. *Journal of Child Language 17*, 85–99.

Hosaka, R., O. Araki, and T. Ikeguchi (2008). STDP provides the substrate for igniting synfire chains by spatiotemporal input patterns. *Neural Computation 20*, 415–435.

Houghton, G. and T. Hartley (1996). Parallels models of serial behaviour: Lashley revisited. *Psyche 2(25)*. Symposium on Implicit Learning.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*. Available online at http://eserver.org/18th/hume-enquiry.html.

Ikegaya, Y., G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste (2004). Synfire chains and cortical songs: Temporal modules of cortical activity. *Science 304*, 559–564.

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Computation 18*, 245–282.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex 17*, 2443–2452.

James, W. (1890). *The Principles of Psychology*. New York: Holt. Available online at http://psychclassics.yorku.ca/James/Principles/.

Jelinek, F. (1990). Self-organized language modeling for speech recognition. In A. Waibel and K.-F. Lee (Eds.), *Readings in Speech Recognition*, pp. 450–506. San Mateo, CA: Morgan Kaufmann. Originally published as an IBM internal report, November 1982.

Jensen, O. and J. E. Lisman (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends in Neurosciences 28*, 67–72.

Ji, D. and M. A. Wilson (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience 10*, 100–107.

Jones, M. W. and M. A. Wilson (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS Biology 3*, 2187–2199.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Kemp, C. and J. B. Tenenbaum (2008). The discovery of structural form. *Proceedings of the National Academy of Science 105*, 10687–10692.

Kersten, D. and A. Yuille (2003). Bayesian models of object perception. *Current Opinion in Neurobiology 13*, 1–9.

Küntay, A. and D. Slobin (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. Slobin and J. Gerhardt (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp*, pp. 265–286. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*, pp. 112–146. New York: Wiley.

Lee, A. K. and M. A. Wilson (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron 36*, 1183–1194.

Levinson, S. C. (2006). On the human interactional engine. In N. Enfield and S. C. Levinson (Eds.), *Roots of Human Sociality*, pp. 39–69. Oxford: Berg.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition 106*, 1126–1177.

Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus 6*, 579–590.

Levy, W. B., A. B. Hocking, and X. Wu (2005). Interpreting hippocampal function as recoding and forecasting. *Neural Networks 18*, 1242–1264.

Lieberman, P. (2002). *Human language and our reptilian brain: the subcortical bases of speech, syntax, and thought*. Cambridge, MA: Harvard University Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Erlbaum. Volume 1: Transcription format and programs. Volume 2: The Database.

Mäkinen, E. and T. Systä (2002). Minimally adequate teacher synthesizes statechart diagrams. *Acta Informatica 38*, 235–259.

Marr, D. and T. Poggio (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull. 15*, 470–488.

Merker, B. (2004). Cortex, countercurrent context, and dimensional integration of lifetime memory. *Cortex 40*, 559–576.

Merker, B. and K. Okanoya (2007). The natural history of human language: Bridging the gaps without magic. In C. Lyon, C. L. Nehaniv, and A. Cangelosi (Eds.), *Emergence of Communication and Language*, pp. 403–420. London: Springer-Verlag.

Meyer, P., A. Mecklinger, T. Grunwald, J. Fell, C. E. Elger, and A. D. Friederici (2005). Language processing within the human medial temporal lobe. *Hippocampus 15*, 451–459.

Müller, R.-A. and S. Basho (2004). Are nonlinguistic functions in 'Broca's area' prerequisites for language acquisition? FMRI findings from an ontogenetic viewpoint. *Brain and Language 89*, 329–336.

Muller, R. U. and J. L. Kubie (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. *Journal of Neuroscience 9*, 4101–4110.

Nelson, K. (1977). Facilitating children's syntax acquisition. *Developmental Psychology 13*, 101–107.

Newmeyer, F. (1998). *Language Form and Language Function*. Cambridge, MA: MIT Press.

Okanoya, K. and B. Merker (2007). Neural substrates for string-context mutual segmentation: A path to human language. In C. Lyon, C. L. Nehaniv, and A. Cangelosi (Eds.), *Emergence of Communication and Language*, pp. 421–434. London: Springer-Verlag.

O'Keefe, J. and J. Dostrovsky (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely moving rat. *Brain Research 34*, 171–175.

Onnis, L., H. R. Waterfall, and S. Edelman (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition 109*, 423–430.

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science 314*, 91–94.

O'Reilly, R. C. and M. J. Frank (2006). Making working memory work: a computational model of learning in the frontal cortex and basal ganglia. *Neural Computation 18*, 283–328.

O'Reilly, R. C. and K. A. Norman (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences 6*, 505–510.

Pereira, A. F., L. B. Smith, and C. Yu (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connection Science 20*, 73–89.

Pereira, F. (2000). Formal grammar and information theory: Together again? *Philosophical Transactions of the Royal Society 358*(1769), 1239–1253.

Peyrache, A., M. Khamassi, K. Benchenane, S. I. Wiener, and F. P. Battaglia (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience 12*, 919–929.

Pickering, M. J. and S. Garrod (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27*, 169–225.

Pickering, M. J. and S. Garrod (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences 11*, 105–110.

Pietroski, P. M. (2003). The character of natural language semantics. In A. Barber (Ed.), *Epistemology of Language*, pp. 217–256. Oxford, UK: Oxford University Press.

Putnam, H. (1967). The 'innateness hypothesis' and explanatory models in linguistics. *Synthese 17*, 12–22.

Quine, W. V. (1961). The problem of meaning in linguistics. In S. Saporta (Ed.), *Psycholinguistics: a book of readings*, pp. 251–261. New York: Holt, Rinehart and Winston.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Richardson, D. C. and R. A. Dale (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science 29*, 39–54.

Richardson, D. C. and M. J. Spivey (2000). Representation, space and Hollywood Squares: looking at things that aren't there anymore. *Cognition 76*, 269–295.

Saffran, J. R., R. N. Aslin, and E. L. Newport (1996). Statistical learning by 8-month-old infants. *Science 274*, 1926–1928.

Schendan, H. E., M. M. Searl, R. J. Melrose, and C. E. Stern (2003). An fMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron 37*, 1013–1025.

Seger, C. A. (2006). The basal ganglia in human learning. *The Neuroscientist 12*, 285–290.

Seger, C. A. and C. M. Cincotta (2006). Dynamics of frontal, striatal, and hippocampal systems in rule learning. *Cerebral Cortex 16*, 1546–1555.

Shapiro, M. (2009). Memory networks: answering the call of the hippocampus. *Current Biology 19*, R329–R330.

Smith, L. B. and M. Gasser (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life 11*, 13–30.

Smith, L. B. and C. Yu (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition 106*, 333–338.

Solan, Z., D. Horn, E. Ruppin, and S. Edelman (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Science 102*, 11629–11634.

Stenning, K. and M. van Lambalgen (2008). *Human reasoning and cognitive science*. Cambridge, MA: MIT Press.

Syal, S. and B. L. Finlay (2009). Motivating language learning: thinking outside the cortex. Manuscript.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory 1*, 113–149.

Thal, D. J. and M. Flores (2001). Development of sentence interpretation strategies by typically developing and late-talking toddlers. *Journal of Child Language 28*, 173–193.

Thomaz, A. L. and C. Breazeal (2008). Experiments in socially guided exploration: lessons learned in building robots that learn with and without human teachers. *Connection Science 20*, 91–110. Special Issue on Social Learning in Embodied Agents.

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex 42*, 480–485.

Von Berger, E., B. Wulfeck, E. Bates, and N. Fink (1996). Developmental changes in real-time sentence processing. *First Language 16*, 193–222.

Wallenstein, G. V., H. Eichenbaum, and M. E. Hasselmo (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences 21*, 317–323.

Waterfall, H. R. (2006). *A little change is a good thing: Feature theory, language acquisition and variation sets*. Ph. D. thesis, University of Chicago.

Waterfall, H. R. (2009). A little change is a good thing: The relation of variation sets to children's noun, verb and verb-frame development. Submitted.

Waterfall, H. R. and S. Edelman (2009). The neglected universals: Learnability constraints and discourse cues. *Behavioral and Brain Sciences 32*, 471–472. A commentary on *Universals and cultural variation in turn-taking in conversation* by Evans & Levinson.

Waterfall, H. R., B. Sandbank, L. Onnis, and S. Edelman (2009). An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language* -, –. In press.

Weber, B., J. Wellmer, M. Reuber, F. Mormann, S. Weis, H. Urbach, J. Ruhlmann, C. E. Elger, and G. Fernandez (2006). Left hippocampal pathology is associated with atypical language lateralization in patients with focal epilepsy. *Brain 129*, 346–351.

Werker, J. F. and H. H. Yeung (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences 9*, 519–527.

Wittgenstein, L. (1958). *Philosophical Investigations* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall. Translated by G. E. M. Anscombe.

Yu, C. and D. Ballard (2007). A unified model of word learning: Integrating statistical and social cues. *Neurocomputing 70*, 2149–2165.