

1

Introduction: Nativism in Linguistic Theory

Clearly human beings have an innate, genetically specified cognitive endowment that allows them to acquire natural language. The precise nature of this endowment is, however, a matter of scientific controversy. A variety of views on this issue have been proposed. We take two positions as representative of the spectrum. The first takes language acquisition and use as mediated primarily by genetically determined language-specific representations and mechanisms. The second regards these processes as largely or entirely the result of domain-general learning procedures.

The debate between these opposing perspectives does not concern the existence of innately specified cognitive capacities. While humans learn languages with a combinatorial syntax, productive morphology, and (in all cases but sign language) phonology, other species do not. Hence, people have a unique, species-specific ability to learn language and process it. What remains in dispute is the nature of this innate ability, and, above all, the extent to which it is a domain-specific linguistic device. This is an empirical question, but there is a dearth of direct evidence about the actual brain and neural processes that support language acquisition. Moreover, invasive experimental work is often impossible for ethical or practical reasons. The problem has frequently been addressed abstractly, through the study of the mathematical and computational processes required to produce the outcome of learning from the data available to the learner. As a result, choosing among competing hypotheses on the basis of tangible experimental or observational evidence is generally not an option.

The concept of innateness is, itself, acutely problematic. It lacks an agreed biological or psychological characterization, and we will avoid it wherever possible. It is instructive to distinguish between innateness as a biological concept

Linguistic Nativism and the Poverty of the Stimulus, by Alexander Clark and Shalom Lappin © 2011 Alexander Clark and Shalom Lappin.

2 Introduction

from the idea of innateness that has figured in the history of philosophy, and we will address this difference in section 1.2. More generally, innateness as a genetic property is notoriously difficult to define, and its use is generally discouraged by biologists. Mameli and Bateson (2006) point out that it conflates a variety of different, often not fully compatible, ideas. These include canalization, genetic determinism, presence from birth, and others.

It is uncontroversial, if obvious, that the environment of the child has an important influence on the linguistic abilities that he/she acquires. Children who are raised in English-speaking homes grow up to speak English, while those in Japanese-speaking families learn Japanese. When a typically developing infant is adopted very early, there is no apparent delay or distortion in the language acquisition process. By contrast, if a child is deprived of language and social interaction in the early years of life, then language does not develop normally, and, in extreme cases, fails to appear at all. It is safe to assume, then, that adult linguistic competence emerges through the interaction between the innate learning ability of the child, and his/her exposure to linguistic data in a social context, primarily through interaction with caregivers, as well as access to ambient adult speech in the environment.

The interesting and important issue in this discussion is whether language learning depends heavily on an ability that is special purpose in character, or whether it is the result of general learning methods that the child applies to other cognitive tasks. It seems clear that general-purpose learning algorithms play some role in certain aspects of the language acquisition task. However, it is far from obvious how domain-specific and general-learning procedures divide this task between them. Linguists have frequently assumed that lexical acquisition, for example, is largely the result of data-driven learning, while other aspects of linguistic knowledge, such as syntax, depend heavily on rich domain-specific mechanisms.

Another long-running debate concerns whether the capacity of adults to speak languages can be properly described as knowledge (Devitt, 2006). This is a philosophical question that falls outside the scope of this study. We do not yet know anything substantive about how learning mechanisms or the products of these mechanisms are represented in the brain. We cannot tell whether they are encoded as propositions in some symbolic system, or are emergent properties of a neural network. We do not yet have the evidence necessary to resolve these sorts of questions, or even to formulate them precisely. The technical term *cognizing* has occasionally been used in place of *knowing*, since *knowledge* of language has different properties from other paradigm cases of knowledge. Unlike the latter, it is not conscious, and the question of epistemic justification does not arise. We will pass over this issue here. It is not relevant to our concerns, and none of the arguments that we develop in this book depend upon it.

The idea of domain specificity is less problematic, and it provides the focus of our interest. At one extreme we have details that are clearly specific to language, such as parts of speech. At the other we have general properties of semantic representation, which seem to be domain general in character. We can distinguish

clearly between semantic concepts such as *agent* and purely syntactic concepts such as *subject*, *noun*, and *noun phrase*, even though systematic relations may connect them. Hierarchical structure offers a less clear-cut case. It is generally considered to be a central element of linguistic description at various levels of representation, but it is arguably present as an organizing principle across a variety of nonlinguistic modes of cognition. There are clearly gray areas where a learning algorithm originally evolved for one purpose might be co-opted for another. Most specific proposals for a domain-specific theory of language acquisition do not allow for this sort of ambiguity. Instead, they posit a set of principles and formal objects that are decidedly language specific in nature.

A related question is whether a phenomenon is species specific. Given that language is restricted to humans, if a property is language specific, then it must be unique to people. Learning mechanisms present in a nonhuman species cannot be language specific.

Humans do exhibit domain-general learning capabilities. They learn skills like chess, which cannot plausibly be attributed to a domain-specific acquisition device. One way to understand the difference between domain-general and domain-specific learning is to consider an idealized form of learning. One of the most general such formulations is Bayesian learning. It abstracts away from computational considerations and considers the optimal use of information to update the knowledge of a situation. On this approach we can achieve a precise characterization of the contribution that domain knowledge makes, in the form of a prior probability distribution. In domain-specific learning, the prior distribution tightly restricts the learner to a small set of hypotheses. The prior knowledge is thus very important to the final learning outcome. By contrast, in domain-general learning, the prior distribution is very general in character. It allows a wide range of possibilities, and the hypothesis on which the learner eventually settles is conditioned largely by the information supplied by the input data. This latter form of learning is sometimes called *empiricist* or *data-driven* learning. Here the learned hypothesis, in this case the grammar of the language, is largely extracted from the dataset through processes of induction.

Language acquisition presents some unusual characteristics, which we will discuss further in the next chapter. First, languages are very complex and hard for adults to learn. Learning a second language as an adult requires a significant commitment of time, and the end result generally falls well short of native proficiency. Second, children learn their first languages without explicit instruction, and with no apparent effort. Third, the information available to the child is fairly limited. He/she hears a random subset of short sentences. The putative difficulty of this learning task is one of the strongest intuitive arguments for linguistic nativism. It has become known as *The Argument from the Poverty of the Stimulus* (APS).

The term *universal grammar* (UG) is problematic in that it is not used in a consistent manner in the linguistics literature. On the standard description of UG, it is the initial state of the language learner. However, it is also used in a number of alternative ways. It can refer to the universal properties of natural languages, the set of principles, formal objects, and operations shared by all

4 Introduction

natural languages. Alternatively, it is interpreted as the class of possible human languages. To avoid equivocation, we will take UG in the sense of the term that seems to us to be standard in current linguistic theory. We intend UG to be the species-specific cognitive mechanism that allows a child to acquire its first language(s). Equivalently, we take it to be the initial state of the language learner, independent of the data to which he/she is exposed in his/her environment. We will pass over the systematic ambiguity between UG taken as the actual initial state of the learner, and UG construed as the theory of this state, as this distinction is not likely to cause confusion here. Given this interpretation of UG, its existence is uncontroversial. The interesting empirical questions turn on its richness, and the extent to which it is domain specific. These are the issues that drive this study.

1.1 Historical Development

Chomsky has been the most prominent advocate of linguistic nativism over the past 50 years, though he has largely resisted the use of this term. His view of universal grammar as the set of innate constraints that a language faculty imposes on the form of possible grammars for natural language has dominated theoretical linguistics during most of this period. To get a clearer idea of what is involved in this notion of the language faculty we will briefly consider the historical development of the connection between UG and language acquisition in Chomsky's work.

Chomsky (1965) argues that, given the relative paucity of primary data and the (putative) fact that statistical methods of induction cannot yield knowledge of syntax, the essential form of any possible grammar of a natural language must be part of the cognitive endowment that humans bring to the language acquisition task. He characterizes UG as containing the following components (p. 31):

- 1 (a) an enumeration of the class $s_1, s_2, \dots$ of possible sentences;
- (b) an enumeration of the class $SD_1, SD_2, \dots$ of possible structural descriptions;
- (c) an enumeration of the class $G_1, G_2, \dots$ of possible generative grammars;
- (d) specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j , for arbitrary i, j ;
- (e) specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number).

1(c) is the hypothesis space of possible grammars for natural languages. 1(a) is the set of strings that each grammar generates. 1(b) is the set of syntactic representations that these grammars assign to the strings that they produce, where this assignment can be a one-to-many relation in which a string receives alternative descriptions. 1(d) is the function that maps a grammar to the set of representations for a string. 1(e) is an evaluation measure that ranks the possible grammars.

Specifically, it determines the most highly valued grammar from among those that generate the same string set.

Chomsky (1965) posits this UG as an innate cognitive module that supports language acquisition. It parses the input stream of primary linguistic data (PLD) into phonetic sequences that comprise distinct sentences, and it defines the hypothesis space of possible grammars with which a child can assign syntactic representations to these strings. In cases where several grammars are compatible with the data, the evaluation measure selects the preferred one.

Chomsky distinguishes between a theory of grammar that is descriptively adequate from one that achieves explanatory adequacy. The former generates and assigns syntactic representations to the sentences of a language in a way that captures their observed structural properties. The latter incorporates an evaluation measure that encodes the function that children apply to select a single grammar from among several incompatible grammars, all of which are descriptively adequate for the data to which the child has been exposed. This notion of explanatory adequacy is formulated in terms of a theory of UG's capacity to account for central aspects of language acquisition.

The evaluation measure in the *Aspects* model of UG is an awkward and problematic device. It is required in order to resolve conflicts among alternative grammars that are compatible with the PLD. However, it is not clear how it can be specified, and what sort of evidence should be invoked to motivate an account of its design. By assumption, it ranks grammars that enjoy the same degree of descriptive adequacy, and so the PLD cannot help with the selection.

Notions of formal simplicity of the sort used to choose among rival scientific theories do not offer an appropriate grammar-ranking procedure for at least two reasons. First, they are notoriously difficult to formulate as global metrics that are both precise and consistent. Second, if one could define a workable simplicity measure of this kind, then it would not be part of a domain-specific UG but an instance of a general principle for deciding among competing theories across cognitive domains. Chomsky (1965, p. 38) suggests that the evaluation measure is a domain-specific simplicity measure internal to UG.

If a particular formulation of (i)–(iv) [1(a)–1(d)] is assumed, and if pairs (D_1, G_1) , (D_2, G_2) ... of primary linguistic data and descriptively adequate grammars are given, the problem of defining “simplicity” is just the problem of discovering how G_i is determined by D_i for each i . Suppose, in other words, that we regard an acquisition model for a language as an input-output device that determines a particular generative grammar as “output,” given certain primary linguistic data as input. A proposed simplicity measure, taken together with a specification (i)–(iv), constitutes a hypothesis concerning the nature of such a device. Choice of a simplicity measure is therefore an empirical matter with empirical consequences.

The problem here is that Chomsky does not indicate the sort of evidence that can be used to evaluate such a simplicity metric. If observable linguistic data and

6 Introduction

general notions of theoretical simplicity are excluded, then we have only the facts of language acquisition to go on. But it is not obvious how these can be used to define a UG internal evaluation function. If, at the final stage of the acquisition process, several descriptively adequate grammars are available for a language L , then how will we know which of these a child's evaluation metric selects as the most highly valued grammar for L ? We seem to be left with a mechanism whose description is inaccessible to the empirical assessment that Chomsky insists is the only basis for understanding its design.

A solution to this problem was proposed with the emergence of the Principles and Parameters (P&P) model of UG. Chomsky (1981) suggests that UG consists of schematic constraints on the representations that comprise the syntactic derivation of a sentence, and on the movement operation which specifies the mappings between adjacent levels in the derivation. These constraints include parameters that allow for a finite number of possible values (ideally they are binary). Assigning values to all the parameters of UG yields a particular grammar.

In the P&P framework, language acquisition is construed as the process of setting parameter values through exposure to a small amount of data from a language. As UG contains a limited number of principles with a bounded set of parameters, each taking a restricted range of possible values, it defines a finite set of possible (core) grammars for natural language. The grammar evaluation measure of the *Aspects* model is no longer needed, and the ranking of competing grammars is dispensed with. Identifying values for the parameterized constraints of UG is intended to yield a unique grammar for the string set of a language, on the basis of the PLD.

Chomsky (1981, pp. 11–12) claims that this finiteness property of the Government and Binding (GB) model of UG “trivializes” important aspects of the computational learning problem of grammar induction. In support of this view he observes that if UG allows for a finite number of grammars, then for any set S of sentences of length k and any possible grammar G , it may be possible to decide, for the elements of S , membership or nonmembership in G , even if G 's full string set is not decidable.

In fact the assertion that having a finite set of possible natural language grammars trivializes the learning problem is inaccurate. As we will see in Chapter 5, a finite hypothesis space of possible grammars is neither a necessary nor a sufficient condition for efficient learning. Grammar induction within a finite hypothesis space can be intractable, while efficient learning in certain types of infinite space is possible. The complexity of the learning process, measured both in terms of the required number of data samples and of the amount of time needed for computation, is a crucial consideration in determining the learnability of a class of languages, even when the string sets of these languages are decidable.¹

The Minimalist Program (MP – Chomsky, 1995, 2001, 2005, 2007) significantly revises the P&P framework. It eliminates intermediate levels of representation (D and S structure in the GB models) in the derivation of a sentence, and so discards the constraints that specify well formedness for these representations. Only the two interface levels of LF (Logical Form, or the Conceptual–Intentional interface) and PF (Phonetic Form or the Sensory Motor

interface) remain, as the outputs of a syntactic derivation from a selection of lexical items (a numeration).

The MP radically simplifies the phrase structure component of the grammar. A single operation, Merge, combines lexically projected functional heads with their complements and their adjuncts to produce a hierarchical tree structure.² Syntactic movement is characterized as “internal merge,” a procedure that copies subconstituents of a tree in a specified position at the left or right edge of the containing phrase structure. Movement is triggered by the need to check features in the lexical head of a constituent against those in the target site. Uninterpretable features must be eliminated through checking prior to the construction of the interface levels, as their presence in such representations will cause an interface to be “illegible” to the cognitive device that it feeds.

Derivations are subject to locality constraints, where these are stated as economy conditions. In earlier formulations of the MP they were expressed as global constraints on entire sets of derivations from a given numeration. In more recent versions, they have been replaced by local economy conditions, intended to serve as restrictions on possible continuations of a derivation from a specified point.

The guiding principle behind the MP is that UG is a “perfect” computational system that provides an optimal mapping from a lexical numeration to the two interface levels of LF and PF. Like the grammar evaluation measure of the *Aspects* model, these notions of perfection and optimality are not characterized independently of the theory of grammar that they are intended to motivate.³ As a result, it is not clear what predictions they make concerning the formal properties of UG, or how to test the comparative “perfection” of alternative theories of UG formulated within the MP.

In contrast to earlier theories of UG, the MP posits a greatly reduced language faculty. In the GB model, parameters are located in general constraints on levels of representation and on movement. In the MP model they have been moved to the functional heads of the lexicon. In fact Boeckx (2008) proposes eliminating parameters from “narrow syntax” entirely and situating them at the PF interface as part of the “Spell Out” of representations at this interface. LF representations are considered to be uniform across languages, and so there is no need to parameterize their mode of realization.

Hauser *et al.* (2002) reduce the “narrow” language faculty (FLN) to recursion, which allows for operations that generate unbounded sets of expressions and hierarchical structures which represent their syntactic form. They suggest that recursion may have first emerged in other cognitive domains, specifically computation with numbers, and then been adapted to language. Pinker and Jackendoff (2005) – among others – observe that the directionality of this purported adaptation is far from obvious. They point out that the elements of recursion, such as hierarchical organization and self-embedding, are pervasive across various primate and human cognitive domains (*inter alia*, navigation, recognition of family and social structures, and perception of geometric mereology). Therefore, it is not clear in which sense recursion can be taken as constitutive of the language faculty.

8 Introduction

Nor is it clear how the highly depleted residue of UG that the MP retains can support the demands of language acquisition that the language faculty was originally proposed to meet. It is a drastic retreat from the richly articulated, domain specific mechanisms specified in Chomsky’s previous theories. Chomsky argued that these elaborate devices were required precisely because domain-general procedures were not adequate to overcome the poverty of PLD available for first language acquisition. The advocates of the MP give no indication of how supplanting a rich language faculty with one that is so impoverished that it fades into an application of principles and procedures shared with other cognitive capacities, will solve the problems for which the earlier models were designed. In fact explaining language acquisition appears to have been inexplicably demoted from the primary objective that a theory of UG is required to satisfy to a peripheral interest of the MP.

In this monograph we will not dwell on the historical development of Chomsky’s theory of UG, although in Chapter 10 we will briefly return to a comparison of the concepts of parameter that the GB and MP models of grammar invoke. Our main interest is to clarify and evaluate the argument from the poverty of stimulus for a domain-specific language faculty. Therefore we will focus on the learning theoretic issues that this argument raises. First, however, we will consider the relation between linguistic nativism and the more general debate between nativists and anti-nativists in cognitive science.

1.2 The Rationalist–Empiricist Debate

In the *Meno* Plato offers one of the first explicitly nativist accounts of human knowledge. Socrates interrogates Menon’s slave boy on the problem of how to construct a square with an area of 8^2 by extending one that is 4^2 . On the basis of the boy’s answers to his questions Socrates eventually guides him to the correct procedure. Socrates concludes that, as the boy had never studied geometry, then he must have been brought to “remember” the geometric principles that he understands. This knowledge had to be inherent within the boy’s soul rather than acquired through learning.⁴

Not only does this section of the *Meno* present an early defence of nativism. It also provides a paradigm of the argument from the poverty of stimulus as the motivation for a nativist claim.

It is important to recognize that the nativist view that Plato is proposing is epistemically normative in character. He is claiming that knowledge, as opposed to opinion and conjecture, cannot be acquired through experience. It can only be achieved through rational reflection and intuition, where the content of knowledge corresponds to propositions that are necessarily true and known to be so through sound methods of reasoning. Therefore, this variety of nativism is not a theory of how human cognition works in the natural world, but a claim about what constitutes knowledge and how it is to be obtained.

The rationalists of the seventeenth and eighteenth centuries also take the identification of reliable foundations for knowledge as their primary concern in developing their respective epistemological theories. They regard experience as incapable of supplying an independently adequate basis for knowledge of the world. Instead they propose to derive it through inference from a small number of propositions grasped as necessarily true through clear rational understanding.⁵

Descartes (1965, p. 94) takes clear and distinct ideas to express the content of true propositions whose certainty is beyond doubt. He finds that the most accessible of such ideas convey knowledge of his own mind. He concludes the Second Meditation with the following observation.

...for, since it is now manifest to me that bodies themselves are not properly perceived by the senses nor by the faculty of the imagination, but the intellect alone; and since they are not perceived because they are seen and touched, but only because they are understood [or rightly comprehended by thought], I readily discover that there is nothing more easily or clearly apprehended than my own mind.

Spinoza (1934) takes mathematical reasoning as the model of a reliable procedure for discovering the essential properties of an object. Knowledge of these depends not on experience of the object, but rational intuition, which supplies the initial premises from which its nature may be deduced.

Leibniz (1969) revises Descartes' condition that an idea is true iff it is clear and distinct by requiring that true ideas be understood *a priori* as possible. By this he seems to intend that we have adequate knowledge of an entity or a phenomenon to the extent that we recognize its constitutive or defining properties.

By contrast, the empiricists of this period seek both to explain the origins of human cognition (ideas), and to evaluate its epistemic status on the basis of these origins. Their first concern is broadly psychological, while their second is epistemic in the sense that preoccupies the rationalists.

We depart from Cowie (1999) in understanding the rationalists to be not primarily interested in the natural origins of cognition, but focused on the discovery of knowledge. She takes the rationalists, as well as the empiricists, to be concerned with both the psychological and the epistemic questions.

It seems to us that the rationalists obtain their theory of knowledge from their respective metaphysical systems. Descartes' dualism, Spinoza's monism, and Leibniz's pluralism of hermetically distinct monads lead each of these thinkers to exclude sense experience as a possible source of genuine knowledge of the essential properties of objects and events in the world. The empiricists adopt the opposite approach. They derive their ontology from their theory of knowledge. If the ideas of sense experience are the foundations of knowledge, then everything that we know (and can know) about the world and the mind must be derived from these ideas and the operations that the mind applies to them.

10 Introduction

We also disagree with Fodor (2000), who regards Chomsky’s nativist view of grammar as a direct descendent of rationalist epistemology. He takes this view to construe grammar as knowledge in an epistemic sense. There are two problems with Fodor’s claim.

First, it is not obvious how grammar can be assimilated to the sort of knowledge with which the rationalists are concerned. The latter consists of propositions about the world that can be demonstrated to be not simply true but certain.

Second, it is by virtue of their status as necessary truths that these propositions can only be known through rational reflection and analysis. By contrast, Chomsky’s nativist assertion of a language faculty is an empirical claim requiring factual support, like any other scientific hypothesis. It is not a first principle of knowledge or metaphysics, but a statement about the relationship between human biology and natural language.

One of the empiricists’ primary psychological interests is to identify the procedures of the mind that generate complex and abstract ideas from simple ideas of sensory experience. Locke (1956, pp. 75–76), for example, posits three such operations.

The acts of the mind wherein it exerts its power over its simple ideas are chiefly these three: (1) Combining several simple ideas into one compound one; and thus all *complex ideas* are made. (2) The second is bringing two ideas, whether simple or complex, together, and setting them by one another, so as to take a view of them at once, without uniting them into one; by which it gets all its *ideas of relations*. (3) The third is separating them from all other ideas that accompany them in their real existence; this is called abstraction: and thus all its *general ideas* are made.

Hume (1888, p. 11) posits the recognition of resemblance, contiguity in time or place, and cause and effect as the three main elements of the mechanism that associates ideas in the mind. While this mechanism explains the processes through which our beliefs about the world are formed, it does not justify or ground these beliefs.

In his discussion of causality Hume argues that the perception of cause and effect reduces to the regular conjunction, in temporal sequence, of two types of ideas. However, previous co-occurrence does not entail a deeper connection between events. Hume (1888, pp. 91–92) offers an inductivist critique of the rationalist view that observed causal relations follow from deeper properties of entities and events, which can only be known through inquiry into their essential natures.

Thus not only our reason fails us in the discovery of the *ultimate connexion* of causes and effects, but even after experience has inform’d us of their *constant conjunction*, ’tis impossible for us to satisfy ourselves by our reason, why we shou’d extend that experience beyond those particular instances, which have fallen under our observation. We suppose, but are never able to prove, that there must be a resemblance betwixt those objects, of which we have had experience, and those which lie beyond the reach of our discovery.

This dual approach to cognition is also evident in modern empiricist work. Quine (1960) presents a narrowly empiricist account of language acquisition that explains learning on the basis of the pairing of utterances with observed events. He then draws skeptical conclusions from this theory concerning what we can know about meaning and grammar. He argues that the semantics of expressions and their syntactic structure suffer from a radical indeterminacy, as there are no facts beyond the patterns of utterances observed in the presence of objects and events (*stimulus meaning*) available to select among competing interpretations and syntactic analysis of these expressions.

In effect, Quine’s rejection of a defined formal syntax for natural language and intensional notions of meaning are directly analogous to Hume’s criticism of essentialist notions of causality inherent in rationalist theories of knowledge. Moreover, like Hume he builds his epistemic argument on a psychological account of the origin of cognition.

Current debates between advocates and critics of cognitive nativism are sometimes described as a continuation of the rationalist–empiricist dispute. Such descriptions misrepresent these debates. The focus of disagreement between rationalists and empiricists is not the source of cognition as such, but its epistemic reliability. Rationalists insist that sensory experience does not provide a solid basis for knowledge, because it is susceptible to uncertainty and confusion. Genuine knowledge can only be achieved through rational intuition, and valid inference from necessary first principles. Empiricists argue that sensory experience constitutes the only source of simple ideas, and all other cognition is generated through combinatory and analytic operations on this input. Hume acknowledges that the ideas obtained in this way do not achieve the knowledge of essential properties of objects that the rationalists seek, but he concludes that such knowledge is not possible.

Neither contemporary nativists nor their critics are seeking to evaluate the reliability of cognition as a source of information about the world. They hold different views on how this information is acquired. Nativist accounts of a given cognitive ability rely heavily on the assumption of an innate, domain-specific device that determines the emergence of cognition in that area. This device is regarded as biologically grounded through encoding in the human genotype. Anti-nativists also posit a rich set of innate learning mechanisms, but these are generally of a domain general character, with application to a variety of cognitive areas.

There is broad agreement between advocates and opponents of cognitive nativism that the issues that divide them are empirical in nature. They concur that these issues can only be decided by scientific investigation of the psychological, neural, and genetic basis for different kinds of human cognitive development.

1.3 Nativism and Cognitive Modularity

An important feature of nativist theories of cognition is the identification of certain cognitive abilities with distinct psychological modules. These are innately determined, task-specific devices for processing input of a particular kind.

12 Introduction

Fodor (1983) proposes an influential version of such a modularized mental architecture. He posits units for each of the five sensory modes, and a language faculty as the primary module. On Fodor’s account a module is “informationally encapsulated.” It handles only input from a specific domain, which it maps into a symbolic (“syntactic”) form that it passes to a central processing component. The latter integrates input from different modules, and performs inferences on the symbolic structures it receives from them. While a module cannot access information from outside of its domain, the central system applies complex learning and reasoning procedures to content from a variety of sources.

Modules perform their processing operations rapidly and automatically, without the intervention of central component reasoning. The language module recognizes phonetic sequences, organizes them as phonological and morphological strings, and parses them into syntactic structures. These are transferred to the central component as lexically filled logical forms, where they are interpreted in conjunction with the symbolic forms received from other modules.

Fodor (2000) claims that while modules process their input by local computational operations, the central component applies global procedures to the multimodal set of symbolic forms that it receives in order to perform holistic, context-dependent inferences like those involved in abduction. He suggests that these global inference patterns, which generate beliefs and other mental states with propositional content, cannot be analyzed by the same methods that have been applied to modular aspects of cognition. He concludes that these higher cognitive functions have so far resisted scientific understanding, and their explanation remains a major challenge for future work in cognitive science.

Fodor is critical of the attempts by neo-Darwinian nativists, such as Pinker (1997a), to assimilate all mental functions, including higher cognitive activity, to a comprehensively modular model. He argues that this “massively modular” approach cannot account for the global nature of conscious mental processes like abduction.⁶

1.4 Connectionism, Nonmodularity, and Antinativism

Antinativists generally reject modularized cognitive architecture, and they argue that task-general learning procedures can account for most cognitive operations. An important challenge to the nativist paradigm comes from the family of approaches broadly called *Connectionist*, which uses neural networks as models of the mind/brain.⁷

Neural networks are simple statistical learning mechanisms that are thought to resemble, in some respects, the neural architecture of the brain. In these models, units corresponding to neurons receive inputs and process them, passing outputs to other units to which they are connected in a network. The final outputs constitute the value that the mechanism generates for a specified set of inputs. Multilayer feed-forward networks are widely used for connectionist modeling. Figure 1.1 shows a schema for this kind of system. The hidden units can modify information

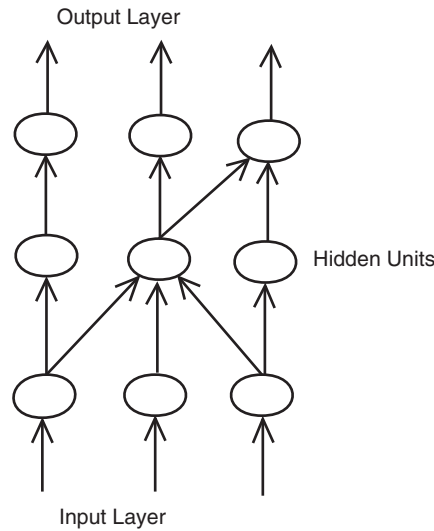


Figure 1.1 Multilayer feed-forward connectionist network with hidden units

received from the initial input units, as elements of a complex function that maps this data to the final output. There is no upper bound on the number of intermediate hidden layers that such a network may contain, or the number of units at any of its levels.

Inputs receive numerical values indicating their strength, and the connections between units are weighted. For any given input I_i transferred along a connection C_i to a unit U_j , an input value IV_i is computed as the product of I_i 's strength and C_i 's weight. U_j may receive inputs from several sources, and its total input $I_{U_j} = \sum_{i=1}^k IV_1 \dots IV_k$. U_j applies a function f_{C_j} to I_{U_j} to produce its output.

A procedure commonly used to train neural networks of this kind is the backpropagation algorithm. The weights of the connections in the network are initially assigned random values. In each training phase the system's output is compared to a target value, and the degree of error, measured as the difference between the actual output and the target, is transferred back from the output units through the network. The backpropagation algorithm computes an error value for each neuron, which represents its contribution to the total error rate of the system. The connections among the units are adjusted to reduce these error values. This process is iterated through successive training cycles in order to improve the network's performance.

Multilevel feed-forward networks using backpropagation have been highly successful at a variety of pattern classification tasks, such as face identification and optical character recognition. One of their important limitations, which restricts their capacity to learn certain complex tasks, is the absence of a memory for encoding previous output values of different units. Elman (1990) proposes a

14 Introduction

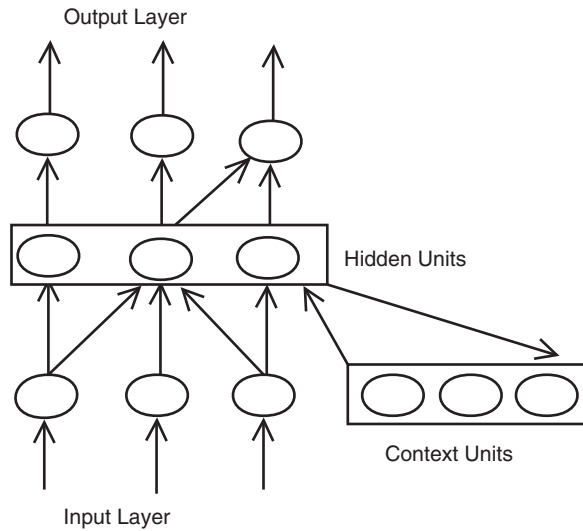


Figure 1.2 Simple recurrent network

simple recurrent network (SRN), which extends multi-level networks by adding a device for recording the previous outputs of the hidden units in a set of context units. The latter feed the outputs of the hidden units back to them in the next activation phase, to enable them to be used in computing the output values of the next set of inputs to the network. Figure 1.2 exhibits the structure of a SRN.

Context units provide SRNs with a set of stacks for storing the immediately preceding environment in which a given input occurs. Elman (1990) uses SRNs to organize lexical items into a hierarchy of semantic classes. Morris *et al.* (1998) show that an SRN can acquire grammatical relations from noun–verb and noun–verb–noun sequences.

Elman (1991, 1998) constructs an SRN that recognizes the sentences of a context-free language fragment of English, which includes transitive and intransitive sentences, and multiply-embedded subject and object relative clauses. It does this by modeling the probability of a word as the continuation of a specified sequence in a test corpus, where this probability value is compared to the actual probability of the expression occurring in that context, given its conditional probabilities of occurrence measured in the training corpus. Cartling (2008) proposes a modified version of Elman’s SRN that improves on its performance for the same training corpus.

Marcus (2001) criticizes connectionist claims to provide a viable nonsymbolic model of learning and cognition that dispenses with rules and symbol manipulation. He argues that neural nets like SRNs, in the types of implementations that Elman and other connectionist theorists describe, do not express abstract

generalized relationships, which he characterizes as relations among variables. He also claims that they do not fully capture the relations among constituents of complex structures of the kind that recursive rules specify.⁸

Are Marcus’ criticisms well motivated? Neural nets, and particularly SRNs, can correctly classify new data that they have not previously encountered. They can process new syntactic structures on which they have not been trained. They are also able to handle subject–verb agreement in complex constructions containing relative clauses embedded several layers down in relative and complement clauses. In this sense these networks can be said to have implicit knowledge of the context-free grammar (CFG) that generates these sentences, where the grammar contains recursive rules.

However, there is a point to Marcus’ objections. The SRNs that Elman, Cartling, and others propose for language acquisition learn to recognize the string set of a language generated by a grammar. However, they do not assign parse structures to the elements of this set. Therefore, they do not explicitly represent syntactic ambiguity, where the same string is assigned competing parses, as in the alternative PP attachments in 2.

- 2 (a) John proved a theorem with a lemma.
- (b) [_S John [_{VP} [_{VP} proved a theorem] [_{PP} with a lemma]]]
- (c) [_S John [_{VP} proved [_{NP} a [_{N'} theorem [_{PP} with a lemma]]]]]

In 2(b) the PP *with a lemma* is an adverb modifying the VP *proved a theorem*. In 2(c) it modifies the head noun *theorem*. Notice that this ambiguity is not lexical. It does not consist in assigning the same lexical item to two distinct classes, and so it cannot be captured through *n*-gram probabilities for distinct word class sequences. Each possible syntactic analysis involves a distinct phrase structure configuration holding among the same set of lexical categories.

One can train an SRN to recognize parse structures of this kind, but the parse trees would have to be presented as target objects during training (Henderson, 2010). This sort of supervised learning cannot reasonably be said to offer a model of human language acquisition. While children do have access to phonetic and lexical strings as PLD, they do not encounter parse structure annotations as elements of this data.⁹

This argument is not entirely decisive in showing that nonsymbolic neural networks are unable to model the acquisition of the full range of grammatical knowledge that is involved in learning a natural language from PLD. It may be possible to represent syntactic ambiguity with an SRN by identifying distinct patterns of activation for N PP and VP PP sequences.

However, even if it is the case that nonsymbolic neural nets are inadequate for expressing important aspects of grammatical knowledge, this result does not entail that human language acquisition requires an elaborate, innately specified, dedicated cognitive module. Neural nets are only one among many machine-learning methods. The debate between strong linguistic nativists and advocates of a largely domain-general approach to language learning does not turn on the

16 Introduction

viability of the connectionist program. As we will see in this study, there are a great variety of nonconnectionist information theoretic learning procedures that can be used for language acquisition tasks. Some of them have produced strikingly impressive results in recent work on unsupervised grammar induction. It is clearly a mistake to try to infer the inadequacy of all empiricist learning models from the flaws of a particular class of connectionist models.

1.5 Adaptation and the Evolution of Natural Language

Linguistic nativism posits a rich, genetically specified UG. This view changes the locus of explanation from language acquisition to language evolution. If we explain the former by invoking a powerful, domain-specific cognitive mechanism, then we commit ourselves to explaining its appearance in the evolutionary processes that produced the human species.

Jackendoff (2008) identifies an Evolutionary Constraint on linguistic theories:

Insofar as linguistic competence is not attainable by apes, the human genome must in relevant respects differ from the chimpanzee genome, and the differences must be the product of biological evolution. The richer Universal Grammar is, the more the burden falls on evolution to account for the genetic differences that make it possible for humans but not apes to acquire language. The Evolutionary Constraint, then, puts a premium on minimizing the number and scope of genetic innovations that make the human language capacity possible – and therefore on minimizing the richness of Universal Grammar.

The situation is made more acute by the comparatively short time during which language has emerged. By contrast, nonnativists do not have this explanatory problem. Given that they rely primarily on domain-general adaptations, evolutionary plausibility is not an issue. On their view, learning processes are adaptations or extensions of antecedent cognitive procedures, and we can find many of these in more primitive form in other species. This affords a much longer time span for their development.

Linguistic nativists are divided on the nature of the evolutionary adaptation that originally produced the language faculty.¹⁰ Chomsky (1995, 2007), Hauser *et al.* (2002), Fitch *et al.* (2005), and Fodor (2000) claim that UG was not selected because of the advantage that it conferred on humans for communication. Chomsky argues that natural language is not well designed for communicative purposes, and he suggests that it emerged through a mutation that modified the architecture of the brain in a way that permitted the generation of internal monologues to facilitate the formulation of intentions and plans. Vocalization of internal semantic representations for purposes of communication developed later as a subsequent extension of the language faculty. Chomsky (2007) summarizes this view as follows:

Generation of expressions to satisfy the semantic interface yields a “language of thought.” If the assumption of asymmetry [between mapping to LF and to PF] is correct, then the earliest stage of language would have been just that: a language of thought, used internally. It has been argued that an independent language of thought must be postulated. I think there are reasons for skepticism, but that would take us too far afield.

These considerations provide a very simple thesis about a core part of the evolution of language, one that has to be assumed at a minimum, so it would seem, by any approach that satisfies the basic empirical requirement of accounting for the fact that the outcome of this process is the shared human property UG. At the minimum, some rewiring of the brain, presumably a small mutation or a by-product of some other change, provided Merge and undeletable EF (unbounded Merge), yielding an infinite range of expressions constituted of LIs [lexical items] (perhaps already available in part at least as conceptual atoms of CI [conceptual intentional] systems), and permitting explosive growth of the capacities of thought, previously restricted to the elementary schemata but now open to elaboration without bounds: perhaps schemata that allowed interpretation of events in terms of categorization by some property (hence predication, once Merge is available), actor–action schemata, and a few others that might well have earlier primate origins. Such change takes place in an individual, not a group. The individual so endowed would have the ability to think, plan, interpret, and so on in new ways, yielding selectional advantages transmitted to offspring, taking over the small breeding group from which we are, it seems, all descended. At some stage modes of externalization were contrived. Insofar as third factor conditions operate, UG would be optimized relative to the CI interface, and the mappings to SM [sensory motor] interface would be the “best possible” way of satisfying the externalization conditions. Any more complex account of the evolution of language would require independent evidence, not easy to come by; and some account is needed for any complication of UG that resists principled explanation. A common assumption of paleoanthropology is that emergence of language led to the “great leap forward” exhibited in the archeological record very recently, and the spread of humans all over the world shortly after, all within an eye-blink in evolutionary time.

By contrast Pinker, Bloom, and Jackendoff (Pinker and Bloom, 1990; Pinker and Jackendoff, 2005; Jackendoff and Pinker, 2005) argue that the language faculty is primarily an adaptation driven by communication. Jackendoff and Pinker point out that if communication was incidental to the evolution of UG, then the fact that sentences are externalized through vocalization or signing would be an unexplained coincidence rather than a central feature shaping the evolution of natural language. Moreover, the social aspect of language acquisition would have a secondary role in the evolutionary process.

Indeed, if language were not designed for communication, the key tenet of Minimalism – that language consists of a mapping from meaning to sound – would not be a “virtual conceptual necessity,” as Chomsky has repeatedly asserted, but an inexplicable coincidence. The only way to make sense of the fact that humans

18 Introduction

are equipped with a way to map between meaning and vocally produced sound is that it allows one person to get a meaning into a second person’s head by making a sound with his or her vocal tract.

We note in addition that the innate aspect of the language faculty is for learning language from the community, not for inventing language. One cannot have inner speech without having words, and words above all are learned. (Jackendoff and Pinker, 2005, p. 225)

It is worth noting that Chomsky’s language of thought proposal for the emergence of the language faculty, at least in the version presented in Chomsky (2007), is not really nonadaptationist, as has frequently been claimed. While it does not take communication to be the primary factor selecting UG and determining its shape, it does propose that the improved cognitive capacities that UG enables confer an adaptational advantage.

It is intriguing that Chomsky presents this view as the simplest possible account of language evolution. In fact, this is only the case if one assumes the MP theory of UG in which it is embedded. Specifically, Chomsky’s suggestion requires the assumption that derivations from lexical numerations to the CI interface are a uniform and defining element of UG, which is a “perfect” computational system for realizing these mappings. Mappings to SM, by contrast, are secondary extensions of this system and variable across languages. It also posits an abstract universal conceptual lexicon that preceded phonological or morphological properties. These assumptions are not in any obvious way simpler than those required by the communication based theory, when considered independently of the MP. Nor are they particularly plausible or straightforward. The MP, on which Chomsky’s view depends, is itself largely devoid of empirical support. It is less equipped to deal with language acquisition than the theories of UG that preceded it, and its coverage of the syntactic properties of natural language falls well short of later versions of GB in many areas.

As interesting as both the internal language of thought and the communication driven proposals for human language evolution are, it is not clear how one could test either of them empirically, even indirectly. Physical evidence like fossils do not distinguish between them, nor is it obvious how one could use current information on the human genotype to decide between these or other possible theories.

There is an alternative approach to language evolution that takes it to be the result of the interaction of several domain-general cognitive capacities rather than the emergence of a distinct faculty. This approach does not posit a specialized cognitive mechanism selected for language use. Instead, natural language as a formal system is itself the locus of adaptation.

In a series of computational modeling experiments Kirby (Kirby, 2001, 2007; Kirby and Hurford, 2002) shows that combinatorial structure and compositional semantics emerge without prior specification through sequences of learning cycles from initially arbitrary associations of a set of signals with a set of meanings. In his *Iterative Learning Model*, at each cycle an agent constructs pairings of meanings (sets of vectors) and signals (strings of letters from an alphabet) from which a

learner must induce rules for these mappings. Communication is not part of the model, and so it plays no role in the evolution of the system.

Kirby identifies two factors that exert competing adaptational pressures. Induction (learning from pairing samples) favours regularity in signal-meaning patterns, and so it promotes compositionality. When occurrences of these patterns are relatively sparse in the data available to the learner, compositionality facilitates learning. By contrast, production is biased to prefer short signals, which are more frequent in the data. The result is a system in which the frequency of a signal is inversely correlated with the regularity of the signal-meaning relation that it exemplifies. Infrequent forms tend to be compositional, while frequently occurring ones are short and irregular. Extrapolating from these models Kirby takes these adaptational pressures to operate directly on natural language as a formal system rather than on the genotype of the biological organisms that use it.

Christiansen and Chater (2008) present a novel version of this approach. They claim that, rather than the brain having evolved a language faculty, natural language has adapted to the general cognitive architecture of the brain to facilitate language acquisition and cultural transmission. They rule out specialized biological evolution to support language by arguing that an assumption that either of the two possible ways in which such evolution could have occurred raises insuperable difficulties (they call this the “logical problem of language evolution”).

If natural language were the product of a biological adaptation of the human organism, then one must explain how a uniform, genetically determined UG could have remained stable in radically distinct language environments. Even if the language faculty evolved prior to widespread human dispersion and gave rise to a single Ur language, the rapid divergence of its descendants would create pressures for evolutionary variation in different speech environments. However, there is no evidence for such a local biological adaptation. All typically developing humans can acquire any natural language to which they are exposed in childhood with an approximately equal level of efficiency.

One might respond to this argument by saying that all linguistic variation is restricted by the conditions of UG. Therefore linguistic change does not require further adaptation, as it takes place in accordance with the abstract universals imposed by the language faculty. As Christiansen and Chater suggest, the problem with this reply is that it renders the adaptationist explanation of UG circular. If UG evolved as a stable genetically determined feature of the brain through adaptation to a linguistic environment, then it is not possible to appeal to the presence of UG in order to explain the universals that define the limits of variation for that environment.

Assume, on the other hand, that natural language is a nonadaptational side effect of other biological changes in the species, and is therefore what Gould and Lewontin (1979) describe as an evolutionary spandrel. If that were the case, the complex combinatorial system specified by UG and the constraints that apply to it would have emerged by chance rather than in response to selectional pressure. The probability of such a random biological event is vanishingly small.

20 Introduction

Christiansen and Chater (2008) argue that the formal properties of natural languages are adapted, through generations of cultural transmission, to the non-linguistic cognitive design of the brain to facilitate acquisition and information processing.

Within this general framework, cross-linguistic universals (to the extent that they can be identified) are emergent properties of the processes through which natural language is modified and adapted over generational cycles of language transmission. In contrast to the UG view, this approach does not treat language universals as hard constraints. Instead, they are defeasible patterns for encoding information. These patterns can be assigned conditional probabilities that depend on the values of several variables representing information theoretic properties of the data from which languages are learned. They will also be partially determined by nonlinguistic procedures for learning and information processing.

As in the case of the internal monologue and the communication-based explanations for the origin of the language faculty, there is no independent evidence for this account of language evolution. Nor is it obvious how we might find such evidence. However, it does offer a credible alternative to the strong linguistic nativist view. Its advocates have also demonstrated, through the implementation of evolutionary models, the computational viability of some of its central claims concerning the emergence of the formal properties of natural language.

1.6 Summary and Conclusions

In this chapter we have tried to situate the APS – and the theory of an innately determined dedicated language faculty that it is designed to motivate – within the history of nativism in cognitive science.

Chomsky characterizes UG as the initial state of the language learner. He takes UG to comprise those universal grammatical systems and constraints that cannot be efficiently learned from PLD. In the *Aspects* model (1965), UG included an evaluation metric for ranking grammars compatible with the same set of data. In the P&P framework (introduced in 1981) this measure is discarded in favour of a UG with parameterized principles, where each parameter has a small range of possible values. Chomsky claims that the finitistic properties of this version of UG effectively solves the computational learning issues of language acquisition. As we will see in later chapters, this is not the case. A finite hypothesis space does not ensure efficient learning, and, conversely, in some cases, learning is tractable in an infinite space. The nature and role of parameters in the MP is unclear. On at least one proposal, they are eliminated entirely from the syntactic component of the grammar. The MP has reduced the learning biases and mechanisms of the language faculty to the point that it is not clear how it can support language acquisition in the way that earlier theories of UG were intended to do.

We have seen that while there is a connection between the rationalist–empiricist debate and the APS, this debate is largely tangential to the controversy over linguistic nativism. The former is concerned primarily with the epistemological

source of reliable knowledge, while the latter focuses on empirical questions about the human capacity to acquire and process natural language.

We observed that cognitive nativism in general, and linguistic nativism in particular, posits a modularized architecture of the mind in which task-specific processing units map input from particular domains into representations at interfaces to a central interpretation system.

Antinativists postulate nonmodular models of cognition in which an integrated system applies domain-general learning and processing procedures across different input domains. Connectionists implement neural networks as computational models of this approach. While some of these networks (in particular, SRNs) can be trained to recognize and generate string sets for CFG subsets of natural languages, it is not obvious that they can learn to produce the parse structures required to represent syntactic ambiguity, except by supervised learning on data annotated with such structures. However, whether or not connectionism proves to be a viable theory of grammar induction does not decide the issue of linguistic nativism. Many nonconnectionist machine learning (ML) methods have been applied to a variety of unsupervised language learning tasks with increasingly impressive results.

We discussed three views of language evolution. Chomsky’s internal monologue proposal suggests that the language faculty is the result of a “rewiring” of the brain that produced a device for mapping conceptual lexical items onto CI interfaces in a way that enhanced the human ability to plan and formulate complex intentions. The externalization of CI representations through vocal and gestural means was a subsequent development, with communication playing no role in the original appearance of the initial computational system.

Communication adaptionists such as Pinker, Bloom, and Jackendoff see the language faculty as an incremental adaptation driven largely by the enhanced communicative capacity that language offers.

Kirby, Hurford, Christiansen, and Chater propose nonnativist accounts on which language is not the product of a specialized biological development. They argue that it is a formal system whose emergence and evolution is itself an adaptation to the cognitive architecture and processing patterns of the brain. Christiansen and Chater point out a number of serious difficulties that both nativist views of language evolution encounter.

In the rest of this study we will focus on the learning theoretic issues raised by the APS. We will not attempt to address the biological or evolutionary aspects of linguistic nativism. We will, for the most part, stay away from notions like innateness, which are notoriously difficult to characterize clearly and consistently in biological terms.¹¹ Instead we will examine some of the claims made on the basis of the APS for strong domain-specific learning priors and biases in theories of language acquisition. We will attempt to clarify the nature of these claims and explore the extent to which the APS motivates them. We will look briefly at some psychological and developmental evidence. However our main concern is with the question of what sort of language-specific biases must be incorporated into a computational learning model in order to achieve efficient learning under

22 Introduction

conditions that correspond to those of human learners. In the next chapter we try to clarify the structure and content of the APS, and we then consider some of the arguments that have been used to support it.

Notes

1. See Lappin and Shieber (2007) for a discussion of this point. Oddly, Chomsky (1981, p. 12) acknowledges that the finiteness assumptions of UG do not show that the issues addressed by computational learning theory applied to grammar induction are “pointless,” but he insists that these issues arise only under “idealizing” moves that abstract away from these assumptions. The computational learning models presented in Chapters 5 and 7 show that this is not the case.
2. See Johnson and Lappin (1999) for a detailed critique of the version of the Minimalist Program presented in Chomsky (1995, 2001).
3. See Johnson and Lappin (1999) and Lappin *et al.* (2000a, 2000b, 2001) for critical discussion of the concept of UG as a perfect computational system.
4. In fact, Socrates’ dialogue with the slave is particularly unconvincing as a case for nativism. He suggests most of the boy’s answers to him, and corrects his errors. If anything, it illustrates the efficacy of negative evidence in facilitating inductive learning.
5. Cowie (1999) presents a detailed and illuminating discussion of the relationship between the rationalist–empiricist debate and the history of nativism. She addresses a variety of issues in the theory of cognition and linguistics, and considers the role that the APS plays in supporting linguistic nativism. We recommend Cowie’s book for a thoughtful treatment of the historical background for many of the problems that we address in this monograph. While we are strongly sympathetic to her criticisms of linguistic nativism, we approach this discussion from a different (and narrower) perspective. Cowie is primarily interested in the philosophical aspects of the general controversy between nativists and antinativists. We focus largely on questions in computational learning theory that the linguistic version of the APS raises.
6. See Barrett (2005) for an attempt to defend massive modularity against Fodor’s critique. Barrett proposes that modules can be conceived on the model of enzymes, which are designed to fit specific chemical substrates, and can accept output from other enzymes as input for processing. While this analogy allows for inter-module communication, it is not clear how it solves the problem of global inference patterns that Fodor treats as the main problem for a fully modular theory of cognition. Unless a multimodal enzyme-like module applies global procedures, then it will not be able to perform abduction. If it does use such procedures, then it is no longer a module in Fodor’s sense, but a central inferencing device. We are grateful to Tzu-wei Hung for bringing Barrett’s article to our attention.
7. Rumelhart *et al.* (1986) and McLelland *et al.* (1986) provide a classic and comprehensive statement of the connectionist research program. See Fodor and Pylyshyn (1988) and Pinker and Prince (1988a) for criticism of this approach.

Elman *et al.* (1996) use connectionist models of learning and cognition to argue against nativist suggestions that a set of innate, strongly task-specific modules are required to explain the development of cognitive capacities like human linguistic competence. See Henderson (2010) for an overview of neural nets and their application in natural language processing (NLP).

8. Marcus also maintains that these systems do not properly distinguish between individuals and kinds. We will limit ourselves here to briefly considering the first two criticisms, as these are most directly relevant to the dispute between linguistic nativists and their opponents.
9. Richard Sproat points out to us that, as children do construct semantic interpretations of strings in the PLD, they have access to the semantic representations of these strings. These representations provide at least indirect information on the syntactic parse structures from which they are derived. This observation is certainly correct. However, as we argue in Chapter 3, it is far from obvious that semantic information can provide an independent basis for syntactic learning. Acquiring semantic representations for strings involves inductive learning that, in turn, raises many of the same issues that the grammar induction task poses.
10. Marcus (2006, 2008) proposes that natural language emerged through a series of incremental (and relatively recent) modifications of other cognitive and physiological systems, which provided the components for a gerrymandered language faculty. He is largely concerned with arguing for the characterization of language as an evolutionary “kludge” rather than speculating on the function that this adaptation serves.
11. See Samuels (2004) and Bateson and Mameli (2007) for discussions of the problems involved in defining innateness as a biological property.