PART I

The Nicod Lectures

Chapter 1
Mental Structure

1.1 Locating the Study of Mental Structure in Cognitive Neuroscience

This book is concerned with exploring human nature in terms of the mental structures that play a role in constituting human experience and human behavior. In order to explain what I mean by "mental structure," it is useful to situate the term within the more general enterprise of cognitive neuroscience.

The leading question of cognitive neuroscience is how the brain works, such that it supports or generates cognition—where by "cognition" I mean an organism's understanding or grasp of the world, and its ability to formulate and execute actions in the world. The neuroscience part of the enterprise includes the study of the physical structure and activity of the brain at all scales, from the inner workings of neurons to the overall organization of brain areas. The cognitive part includes characterizing the functional or computational character of mental activity, as well as the organism's phenomenology—how the organism experiences the world. I will use the term *brain* in the customary way to describe the physical body part which accomplishes cognition, and which is the proper domain of neuroscience. I will use the term *mind* to denote the brain seen from the point of view of its functional or computational aspect, and *mind/brain* when I wish to be neutral between the two.

An important goal of the enterprise is to figure out how the functional domain is instantiated in the neural domain—to use a now somewhat outdated analogy, how the brain's software runs on the hardware—and also to figure out how the neural and computational structures support conscious experience. At the moment, this goal seems far off. We know many details of how brain function is localized and many details of how individual neurons and small clusters of neurons function. But I think it is likely to be a long time before we understand how the neurons actually

accomplish anything as complex as, say, language perception or the storage of vocabulary—in detail or even in principle. So the flood of recent advances in understanding the brain by no means undermines studies of the mind. Part of the burden of this book is to emphasize the value of investigating cognition in terms of mental structure.

Cutting across this dimension of the enterprise are developmental questions, at two scales. First, at the scale of the individual: how do the brain, mental functioning, and phenomenology develop in the individual from conception to death? And second, at the scale of evolution: how do characteristics of the species develop over evolutionary time under the pressures of natural selection? The latter question adds to the mix the fascinating issue of interspecies comparison.

Cutting across both these dimensions is how the functions of the mind/brain divide into capacities or domains or modules or faculties, whatever you wish to call them. On one hand, there is a "vertical" division more or less by subject matter: vision, audition, proprioception (the sense of body position and movement), motor control, language, and so forth. And on the other hand, cutting across *this* is a "horizontal" division into, on one hand, the study of mental structure, and on the other, the kinds of machinery that process mental structures, such as working memory, long-term memory, attention, and learning, all of which are involved in each of the "vertical" capacities. Table 1.1 sums up all the dimensions of the inquiry.

Of course, we often study an individual cell in this four-dimensional matrix as though it were isolated—say, the brain localization of some aspect of visual working memory. However, we should understand that the essence of the enterprise lies in characterizing the interaction of these systems.

It is my impression that of all the cognitive sciences, only linguistics has systematically and explicitly investigated the content of mental structures that underlie a human capacity. The rest of cognitive neuroscience has for the most part made do with relatively rudimentary notions of mental structure, exploring more intensely issues of neural localization and/or the "horizontal" capacities of working memory, attention, learning, and the like. Three exceptions: Marr 1982 is the inception of a detailed study of the mental structures involved in vision (with Biederman 1987 as a related endeavor); this style of investigation has receded since Marr's death. Lerdahl and Jackendoff 1983 applies the approach of linguistic theory to music cognition. Finally, chapter 4 compares language with the capacity for complex action.

Table 1.1
Ways of studying the mind/brain

Dimension 1
Neuroscience (brain) vs.
Cognitive science (mind/functional properties) vs.
Behavior and phenomenology
—*plus* relations among the three

Dimension 2
Steady state vs.
Individual development vs.
Evolutionary development

Dimension 3 ("Vertical" capacities or "modules")
Vision vs.
Language vs.
Motor control vs.
Abstract thought vs.
. . .

Dimension 4 ("Horizontal" division; applies to all "vertical" capacities)
Data structures (mental structures) vs.
Processing capacities
   Working memory vs.
   Long-term memory vs.
   Attention vs.
   Learning

1.2   Mental "Structure" versus Mental "Representation"

Since the early days of cognitive science, the term of art for the computational structures in terms of which the mind operates has been "mental representations" or "symbolic representations." The subtitle of this book deliberately substitutes "mental structures"; let me explain why. The structures that a linguist writes on the page, say syntactic trees, are intended as representations of what is in the mind. However, I would maintain that what is in the mind is best not thought of as a representation or a symbol of anything. The reason is that the words "representation" and "symbol" imply an interpreter or perceiver: it is not just that *this* represents or symbolizes *that*, but implicitly that *this* represents or symbolizes *that* **to so-and-so**. But a person in whose mind syntactic structures reside does not perceive them; rather, the person perceives a linguistic utterance by virtue of having these structures in his or her mind. The only thing

that "perceives" syntactic structures is the faculties of mind that process and store syntactic structures, and in fact the term "perceive" is itself suspect in this context.

If we are to take seriously the relation between mind and brain, this is the only possible view of mental structures. The neurons deep inside the brain that are responsible for cognition have no privileged access to the "real world"; they interact only with other neurons. Contact with the "real world" is established only through long chains of connection leading eventually to sensory and motor neurons. If this is the hardware on which mental capacities "run," then mental capacities too are necessarily limited in their contact with the "real world." They are sensitive to the outside environment only insofar as they are connected through functional (or computational) links to the sensory and motor capacities.

In short, I wish to reject all talk of the "intentionality of mental representations," the idea that mental structures are "about" the world in some direct sense. This goes against the grain of much influential philosophy of cognitive science (e.g. Searle 1980; Fodor 1987).[1] The reader is free to understand such rejection in either of two ways. The weaker stance is methodological: even if mental structures are ultimately connected directly to the world by intentionality, there remains the empirical enterprise of characterizing them for their own sake. Taking this stance, we are choosing to study mental structures as a kind of "engineering," temporarily leaving philosophical concerns behind.

The stronger stance is to take the rejection of intentionality as principled—to claim that once the mental structures are properly characterized, there will be no need for a supervenient intentionality. Such a stance fits far more comfortably with the neuroscience. On the other hand, it depends on a promissory note to the effect that *someday* all the problems associated with intentionality will be worked out. But of course we adopt such promissory notes all the time in science. In particular, any sort of materialist philosophy of mind (i.e. any sort of modern cognitive science) takes for granted the promissory note that someday we will be able to relate all mental processes to brain processes.

---

1. For extended discussion of why I reject intentionality, see Jackendoff 1987, chap. 7; 1992a, chap. 8; 2002a, chaps. 9, 10. Some of the more confrontational commentaries on Jackendoff 2002a (e.g. Adams 2003; Higginbotham 2003; Gross 2005; Rey 2006) reflect the degree to which intentionality is still taken as a sine qua non of theories of mental representation.

For the working scientist, the choice between the methodological and the principled stance rarely affects one's work one way or the other. As far as I can see, the main thing that cripples inquiry is to proclaim that without an account of intentionality, all research on mental function is pointless, and to demand that intentionality be explained before any further work proceeds.
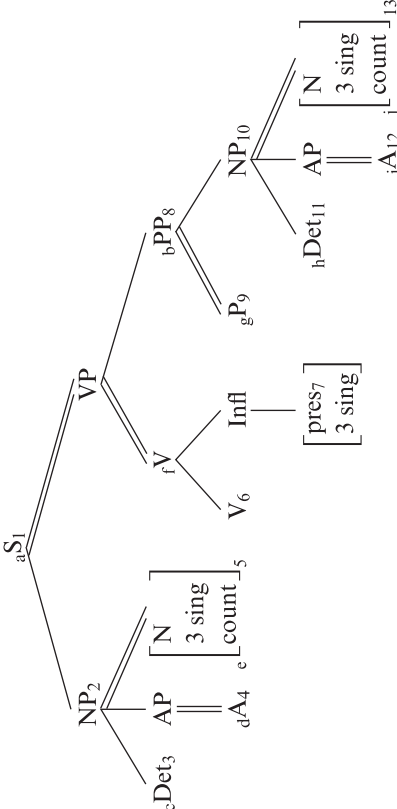
1.3   The Mental Structures of a Simple Sentence

This section presents a very elementary example of linguistic structures as linguists understand them; the next section briefly discusses the issues that such structures raise for neuroscience. Section 1.5 sketches an overall view of the character of the mind in these terms.

So consider someone saying an absolutely simple sentence such as *The little star's beside a big star*. This is quite likely a sentence the speaker has never uttered or heard before. The speaker has constructed it to suit some present communicative context, using elements from his or her long-term memory, in particular the words and the means of putting them together into sentences (the latter often called "rules of grammar"). Linguistic theory is primarily concerned with how words and the principles for combining them are to be characterized *functionally*—as mental data structures, so to speak.
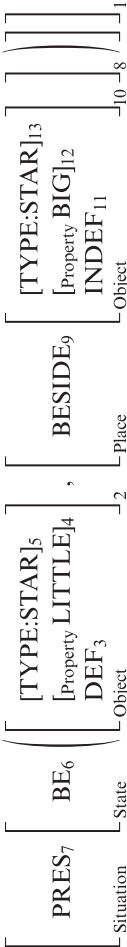
Figure 1.1 (pp. 8–9) shows some of the more prominent aspects of the structure of the sentence *The little star's beside a big star*. These are aspects on which there is substantial agreement among linguists, whatever their creed (Chomskyan or not); there are many disagreements about what further complexity there might be, but there is at least this much. Let me give a brief tour of this structure. (There is more detail in chapter 2, and especially in Jackendoff 2002a, chaps. 1 and 5.)

The upper part of the figure works out the phonological (or sound) structure of the sentence. The basic pronunciation of the sentence appears on the line labeled "segmental structure"; each of the symbols in this line stands for a speech sound. There is substantial agreement that segmental structure is more articulated than this: each speech sound is actually a composite of *phonological distinctive features*. Figure 1.2 (p. 10) shows the decomposition of this level for just the word *star*; you can imagine extending this analysis to the rest of the sentence. The distinctive features capture the dimensions of variation among speech sounds, for instance the position of the tongue, jaw, lips, and velum, and the presence or absence of vocal cord vibration.

Figure 1.1
Structure of *The little star's beside a big star*

Syntactic structure

$_a$S$_1$
NP$_2$ VP
$_c$Det$_3$ AP
$_d$A$_4$ $\left[\begin{array}{l} \text{N} \\ \text{3 sing} \\ \text{count} \end{array}\right]_5$ $_e$

V
$_f$V V$_6$ Infl
$\left[\begin{array}{l} \text{pres}_7 \\ \text{3 sing} \end{array}\right]$

$_b$PP$_8$
$_g$P$_9$ NP$_{10}$
$_h$Det$_{11}$ AP
$_i$A$_{12}$ $_j$ $\left[\begin{array}{l} \text{N} \\ \text{3 sing} \\ \text{count} \end{array}\right]_{13}$

Semantic/conceptual structure

$$\left[\begin{array}{l} \text{PRES}_7 \left[\begin{array}{l} \text{BE}_6 \left(\left[\begin{array}{l} \text{[TYPE:STAR]}_5 \\ \text{[Property LITTLE]}_4 \\ \text{DEF}_3 \end{array}\right]_{\text{Object}}, \left[\begin{array}{l} \text{BESIDE}_9 \left[\begin{array}{l} \text{[TYPE:STAR]}_{13} \\ \text{[Property BIG]}_{12} \\ \text{INDEF}_{11} \end{array}\right]_{\text{Object}} \right]_{\text{Place}} \right)_2 \end{array}\right]_{\text{State}} \end{array}\right]_{\text{Situation}}_1$$

Spatial structure

$_2$ $_{10}$ $_8$

Figure 1.1
(continued)

$$
\begin{bmatrix}
+\text{consonantal} \\
-\text{vocalic} \\
-\text{sonorant} \\
-\text{nasal} \\
+\text{continuant} \\
-\text{voiced} \\
+\text{anterior} \\
+\text{coronal}
\end{bmatrix}
\begin{bmatrix}
+\text{consonantal} \\
-\text{vocalic} \\
-\text{sonorant} \\
-\text{nasal} \\
-\text{continuant} \\
-\text{voiced} \\
+\text{anterior} \\
+\text{coronal}
\end{bmatrix}
\begin{bmatrix}
-\text{consonantal} \\
+\text{vocalic} \\
+\text{sonorant} \\
-\text{nasal} \\
-\text{high} \\
+\text{low} \\
-\text{front} \\
-\text{round}
\end{bmatrix}
\begin{bmatrix}
+\text{consonantal} \\
+\text{vocalic} \\
+\text{sonorant} \\
-\text{nasal} \\
+\text{continuant} \\
+\text{voiced} \\
-\text{anterior} \\
+\text{coronal}
\end{bmatrix}
$$

Figure 1.2
Detail of segmental structure of *star*

Next let's return to figure 1.1. Above the segmental structure is a se-
quence of little tree structures that show how the speech sounds are col-
lected into syllables (notated as σ in the trees). Each syllable contains a
syllabic nucleus (N) and sometimes an onset (O) and coda (C). The nu-
cleus and coda together form the rhyme (R), the part of the syllable that
is used in determining rhyme, and also the part of the syllable that is rel-
evant for determining stress.

Above the syllabic structure is a *metrical grid* of *x*s that marks the
relative stress of the syllables in the sentence: more *x*s above a syllable
indicate more stress. Thus the word *the* is relatively unstressed, and the
word *big* has the maximal stress in the sentence. In turn, the metrical
grid is bracketed into units that represent the *prosodic contours* of the
utterance—its division into breath groups over which intonational con-
tours are defined. In figure 1.1, the bracketing indicates a division some-
thing like *The LITTLE star's—beside a BIG star*. I have not indicated
here the intonation contours themselves; in a tone language such as Man-
darin, there would be additional structure indicating the tones associated
with each syllable.[2]

So far this is just a structured string of sounds; I've said nothing about
the division of the string of sounds into words! This division appears be-
low the segmental structure as another sequence of trees (which for con-
venience are notated upside down), the *morphophonology*. These trees say
that the sentence has five full phonological words: *little*, *star*, *beside*, *big*,
and *star*. Attached to some of them are clitics, corresponding to *the*, *'s*,

---

2. Influential treatments of intonation contours include Pierrehumbert 1980,
Beckman and Pierrehumbert 1986, and Ladd 1996; for tone languages, see Yip
1995. More generally for phonology, see Goldsmith 1995.

and *a*. Notice that the syllabic structure and the morphophonology don't match up exactly. In particular, the clitic *'s* forms part of a syllabic coda with the last consonant of *star.*

All this structure so far is phonology. It says nothing about parts of speech such as nouns and verbs. These categories appear in *syntactic structure*, the next major part of figure 1.1. I have notated this as a tree structure of more or less the familiar sort. There is one important difference: the words *the*, *little*, *star*, and so forth are not notated in the syntactic tree in the conventional fashion. My reason for doing it this way is developed in detail in chapter 2. For now, the basic point is to segregate the different kinds of linguistic features into their proper structures. In particular, the fact that the word is pronounced *star* is a fact of phonology, not of syntax. All the *syntax* knows is that it is a noun, indistinguishable from every other singular count noun in English (in languages such as French, Russian, and Hebrew, grammatical gender would also be notated here).

However, the overall structure must of course indicate that the phonological piece *star* corresponds to a noun in syntactic structure; this is notated with the letter subscripts in figure 1.1. For instance, the subscript e connects the word *star* in morphophonology with the first noun in the syntax. Look also at the clitic z next to *star*, with the subscript f, which is linked to the inflected verb of the sentence. This little z is thus the phonological encoding of the verb *be* in present tense, inflected for third person singular—in other words, the contracted form of *is.*

We've still said nothing about what the sentence means. This is the role of the two structures at the bottom of figure 1.1. The *semantic/conceptual structure* is an algebraic encoding of the propositional organization of the sentence, in function-argument form—a predicate calculus sort of structure. It's over this structure that principles of inference, reference, and truth-conditions can be defined formally. In this particular example, there is a Situation in the present, which consists of a State of a Thing being in a Place. The Thing is of the category STAR, it has the property of being LITTLE, and it is definite (i.e. the speaker takes it to be independently identifiable by the hearer of the utterance). The Place (where the little star is) is a region of space that is determined by a spatial relation, BE-SIDE, in relation to a reference object. In turn, the reference object is also of the category STAR, it has the property of being BIG, and it is indefinite—that is, it is an entity new to the discourse. (For a little more detail, see section 6.1.)

These pieces of the semantic structure are coindexed with the syntactic structure (and therefore indirectly with the phonology) by number subscripts. For instance, the syntactic subject of the sentence (the first NP) has the index *2*, which corresponds with the first Thing constituent of the semantic structure (i.e. the meaning of the phrase *the little star*). Now notice one particular curious correspondence: the semantic feature PRES (present time) has subscript *7*, so it corresponds to present tense in syntax, a feature of the verb's inflection. But this feature of the verb doesn't correspond directly to anything in phonology. Rather, it is swallowed up as part of the inflected verb, which in turn surfaces as the clitic *'s* in phonology—not even a syllabic coda on its own. Thus the outermost functional element in meaning, the one that provides the whole framework for the meaning, surfaces as only a tiny part of the tiniest part of the phonology. This sort of mismatch turns out not to be so unusual in language.

The semantic/conceptual structure in turn maps in some ill-understood way into a spatial or visual encoding of the scene that the sentence describes, so that the sentence can be used to describe a visual scene. I have notated this crudely as the *spatial structure* in figure 1.1 (one could think of this as the "mental model" of the sentence in Johnson-Laird's (1983) sense, or alternatively as a visual percept or visual image). Here the subscripts connect the parts of the visual figure to their corresponding elements in semantic/conceptual structure. The dashed oval in spatial structure corresponds to the spatial region expressed as *beside the star*— something that is not present in visual phenomenology but *is* present in visual understanding. (Of course, in a sentence expressing an abstract proposition, there will be no corresponding spatial structure.)

This completes our tour of the structure of this ridiculously simple sentence. For more complex sentences like those we use constantly, there will be much more of the same. I want to emphasize that all this structure represents a pretty fair consensus among linguists, based on research on thousands of linguistic phenomena in hundreds of languages of the world. This research includes not only speakers' judgments of grammaticality but also analysis of texts, historical change in languages, experimental psycholinguistic research on online processing in perception and production, the acquisition of language by children and adults, the loss of language by aphasics, and so on. I stress the motivation for the analysis because people outside of linguistics sometimes think that linguists just make all this up. Nothing could be farther from the truth: it's the outcome of rigorous empirical research.

1.4   Relevance to Neuroscience

But what does this structure mean—or what *should* it mean—to a neuro-scientist? Of course, there are no symbols like NP and σ running around in our heads. Rather, I think the proper way to understand figure 1.1 is as a claim that there are *functional equivalents* of every element of this structure in our heads. Because this sentence is being produced or understood online, the functional equivalents of these structures must be present in both the speaker's and the hearer's working memory. A sentence is not just a string of words, each of them being a node in a semantic network or some such. It is a set of three or more correlated structures: phonology, syntax, semantics, and (sometimes) spatial structure, each of which has its own particular dimensions of variation, its own repertoire of basic elements, and its own principles of combination. In producing a sentence, one must map from a semantic structure (the meaning one wishes to express), through syntax, to phonology, which leads to the formation of instructions to the vocal tract. In hearing and understanding a sentence, one must convert an acoustic signal into phonology, which in turn can be mapped to syntactic and semantic structures in working memory.[3] Language processing cannot go directly from acoustics to meaning or from meaning to motor control, because the correspondence is determined by the principles of the language: think again of how the meaning 'present time' is related to phonological expression only as a part of the meaning of the little sound z. And in the course of producing or understanding the sentence, the speaker and hearer need all these structures to be available simultaneously in working memory, as is clear from the fact that they know which words correspond to which parts of the meaning.

Naturally, both neuroscientists and linguists would love to know how these structures are instantiated in neural tissue and neural activity. But this is not a question that can be answered at present. In particular, even if we know *where* a structure is localized in the brain—the sort of information that neural imaging can provide—we do not know *how* the brain instantiates the structure. I think it is worth emphasizing our extreme ignorance here. We don't have the slightest idea how even the most elementary units of linguistic structure such as speech sounds can be instantiated

---

3. This is an oversimplification, of course. It is not as though one hears a whole sentence, then parses it all syntactically, then decides what it means. Rather, processing is incremental and involves feedback. See section 1.5.2.

neurally: how speech sounds are stored and how they are processed. Some neuroscientists say we are beyond this stage of inquiry, that we don't need to talk about "symbols in the head" anymore. I firmly disagree. We *know* that language is organized into speech sounds and that speech sounds are only the first step in analyzing linguistic structure. As far as I know, there exist absolutely no attempts to account for even this trivial degree of linguistic complexity in neural terms, and speech sounds only scratch the surface. In my opinion, it is the height of scientific irresponsibility to totally dismiss linguistic theory, claiming that some toy system (say a computational neural network) will eventually scale up to the full complexity of language.[4] A linguist who made comparably ignorant claims about the brain would be a laughingstock. End of sermon.

The structure in figure 1.1 tells us still more about how the brain has to be functionally organized. First, consider the subscripting that connects the structures to each other. This presents an especially complex example of the familiar *binding problem* in neuroscience, a term usually applied to the problem of connecting different aspects of visual representations such as motion, color, and shape, which are (I gather) processed in different brain areas (Treisman 1988). Figure 1.1 shows how to connect different aspects of linguistic representations: sound, grammatical structure, and meaning. What is striking here is that this trivial little sentence requires a staggering amount of binding: each of the 23 subscripts represents a different pair of pieces that has to be connected—simultaneously. Any moderately complex sentence, such as the one you are now reading, requires vastly more binding. This presents a challenge. I gather that the most popular hypothesis for binding is that bound constituents fire in temporal synchrony with each other and out of synchrony with other elements (e.g. Gray et al. 1989; Crick and Koch 1990; Singer et al. 1997). But can this account cope with the binding in figure 1.1? Could there be enough temporal bandwidth in neural firing to discriminate 23 separate bindings at once? This problem is not particular to language, of course; similar problems of massive binding will arise for the integration of any visual scene of medium complexity.

A different challenge for binding arises from the fact that there are two occurrences of the word *star* in the sentence in figure 1.1. Presumably, long-term memory contains one copy of this word. Yet the word *star* must be bound (or copied) to two separate locations in working memory,

---

4. This is approximately the gist of Elman et al. 1996 and Deacon 1997, for instance.

and both copies must be simultaneously present and active in working memory in order for the sentence to be produced, understood, and connected with the visual percepts. But the two copies had better not be bound together; if they were, we would understand there to be a single star that is both little and big! Again, this is not a particularly linguistic problem; it arises any time there are two tokens of the same category in a visual configuration, for instance two identical coins on a table. But the linguistic case points up the essential nature of the problem. The problem, of course, is the necessity for mental representation to be able to discriminate types, stored in long-term memory, from tokens, instantiated in working memory. This is an issue discussed at length by Marcus (2001) in his critique of the most popular variety of connectionist learning.

These issues are treated in more detail in Jackendoff 2002a, especially chapter 3. The message to take from the present discussion is that an investigation of mental structures provides important boundary conditions on the theory of brain function. A similar point was made by Marr (1982) in connection with vision. In both language and vision, if we want to figure out how the brain works, it behooves us to try to understand what functions the mind has to compute. A proposed theory of neural behavior is incomplete if it does not offer genuine solutions to the problems of combinatoriality, structural hierarchy, and binding among structures.

It is not that these problems are particular to language. It is just that linguistic theory focuses on these problems and builds on them in a way that theories of other ''vertical'' faculties of mind usually have not. Part of the message of this book is that these properties recur in other faculties, should we care to look for them. Sixty years ago, nearly everyone thought that language was perfectly transparent and hardly complex at all (and many nonlinguists, even some in psychology and neuroscience, still think so). Since then we have learned that not only is language far more complex than we ever would have dreamed, but so is every other aspect of the mind/brain that has been investigated.

To sum up: Pretty much all cognitive neuroscientists agree in rejecting dualism; ultimately the mind must run in the brain, and there are no mental properties that are causally independent of brain events. However, to insist that neural accounts have absolute priority, that they somehow have a greater reality or are ''more scientific'' than functional accounts, to me has a chilling effect on inquiry. It seems to me that in the practice of research, the relationship between neural and functional accounts ought to be a two-way street: what we know about each dimension of the problem ought to enrich our study of the other.

1.5   An Overall Vision of Mental Architecture

1.5.1   Levels of Structure and Interfaces

Extrapolating from linguistic theory, a vision of an overall "vertical" architecture of mind emerges. The division of the mind into "faculties" and their "subfaculties" is instantiated by a collection of discrete levels of structure, of which phonology, syntax, conceptual structure, and spatial structure are "subfaculties" involved in the language faculty. Each of these levels has its own characteristic basic elements (e.g. distinctive features and syllabic units in phonology) and its own characteristic combinatorial principles (e.g., in phonology, collection of features into segments and concatenation of segments into syllables). In addition, mental structure is governed by *interface* principles that connect particular pairs of levels (or perhaps larger *n*-tuples of levels). Such principles connecting levels L_1 and L_2 establish which parts of L_1 correspond to which parts of L_2; the corresponding parts are bound, as indicated by the subscripts in figure 1.1. A leading question of cognitive science therefore ought to be this:

· What are the levels of mental structure, and what are the interfaces among them?

Notice next that the levels of conceptual structure and spatial structure do not belong to the language faculty per se: they play a role in many different faculties, including vision and action. In contrast, phonology and syntax *are* specific to the language faculty and therefore might be considered (part of) the "narrow language faculty" in the sense of Hauser, Chomsky, and Fitch 2002.[5] The interfaces through which the "narrow language faculty" communicates with conceptual structure and spatial structure are qualitatively not unlike the interface between phonology and syntax: in each case, the interface establishes a correlation between parts of structures.

More broadly, the question arises of how one can talk about what one sees: how the visual faculty communicates with the language faculty. The answer is that the visual faculty comprises a collection of levels connected by interfaces, of which the most peripheral are the distinctions made by the retina and primary visual cortex, and among the most central is the

---

5. Though Hauser, Chomsky, and Fitch do not consider phonology part of the narrow faculty, as they make clear in Fitch, Hauser, and Chomsky 2005, in response to Pinker and Jackendoff 2005.

level of spatial structure. In turn, spatial structure has interfaces that lead into the language faculty. In other words, multimodal interactions are made possible by interfaces that link levels used by the different faculties.

Looking at all the levels "horizontally," we might notice that many different levels of structure are hierarchical, in that elements of structure are combined to make higher-order elements, which in turn combine with other elements. We might further notice that some levels are recursive, in the sense that a structural element of a particular type can form a constituent of another element of the same type. For example, syntactic structure is recursive, in that an element of the type Noun Phrase can be a constituent of another Noun Phrase, as in *[$_{NP}$ the king of [$_{NP}$ the Cannibal Islands]]*, and this embedding can be repeated, as in *[the tip of [the nose of [the father of [the bride of [the king of [the Cannibal Islands]]]]]]*. On the other hand, syllabic structure, though hierarchical, is not recursive, in that such unrestricted embedding is not possible. So a more general question arises:

· Which levels of structure are hierarchical, and, among those, which are recursive?

Of course, this question cannot be answered in a principled way until we have accounts of numerous levels of structure in different faculties. Hauser, Chomsky, and Fitch (2002) speculate that recursion may be unique to humans and in particular to the language faculty, perhaps even *the* single factor that makes language a human specialization; then they back off and speculate that recursion might be found elsewhere in cognition. Jackendoff and Pinker (2005) confirm this speculation, pointing out that figure 1.3 shows evidence of recursion in visual cognition. This display is perceived as being built recursively out of discrete elements that combine to form larger discrete constituents: pairs of *x*s, clusters of four pairs, squares of four clusters, arrays of four squares, arrays of four arrays, and so on. One could further combine four of these superarrays into a still larger array, and continue the process indefinitely. So, to use Chomsky's term, we have here a domain of "discrete infinity" in visual perception, with hierarchical structure of unlimited depth, its organization in this case governed by classical Gestalt principles. Presumably the principles that organize figure 1.3 play a role in perceiving objects in terms of larger groupings, and in segregating individual objects into parts, parts of parts, and so on. Similar principles of grouping appear in music (Lerdahl and Jackendoff 1983).

```
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX

XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX


XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX

XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX



XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX

XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX


XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX

XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
XX  XX   XX  XX      XX  XX   XX  XX       XX  XX   XX  XX      XX  XX   XX  XX
```

Figure 1.3
Recursion in visual grouping

When the similarities and the differences are examined, it appears that hierarchical phrase structure in language cannot be reduced to the principles governing visual and musical grouping. Two formal properties distinguish recursion in syntax. First, elements and phrases of syntax belong to distinguishable syntactic categories such as N or VP; visual groups do not obligatorily fall into some small set of distinguishable categories (as far as we know). The particular family of categories in syntactic phrases appears to be sui generis to syntax. Second, unlike what we find in visual grouping, one member of each syntactic constituent has a distinguished status as *head*, such that the other members are considered dependent on it. Headed hierarchies are found elsewhere in cognition, for instance in syllabic structure (which, as mentioned, is not recursive in the strong sense), in conceptual structure, in certain aspects of musical structures (Lerdahl and Jackendoff 1983; Jackendoff 1987, 249–251), and, as I will argue in chapter 4, in the structure of complex action.

1.5.2   Processing

A theory of structure alone is not a theory of mental functioning. It must be complemented with a theory of how mental structures are processed over time to produce behavior, knowledge, and experience. In present terms, the basic processing operations are the construction of mental structures at each level and the linking of structures at multiple levels. Consider for example language perception. Environmental input leads to the construction[6] of an auditory structure. The interface that links this to phonology leads to the construction of a candidate phonological structure in working memory. The further interfaces to syntax and thence to conceptual structure lead to construction of an interpretation for the heard utterance, also in working memory.[7] However, this process of construction is not just an autonomous function of working memory. In order to get from a phonological string to a meaning, the processor must call on material stored in long-term memory. In particular, the words of the utterance must be identified in order to assign particular chunks of phonology to chunks of meaning. If the hearer doesn't know the words, the meaning cannot be determined. If language perception is successful, the outcome is a set of structures linked in working memory (where the linkings are notated in figure 1.1 as the matched subscripts). In turn, one or more of these structures may be shipped to long-term memory—if only the conceptual structure, one remembers the gist; if all the structures, one has memorized the sentence.

Pretty much everyone imagines similar processes of construction emerging in visual perception, with structure propagating up from sensory to central levels. The main doubt that might arise is whether visual

---

6. Or in this particular case, "transduction," in the sense of Pylyshyn 1984. The process by which sensory stimulation gives rise to a functional organization in the mind cannot be characterized in functional terms—only its output can. In other words, this marks the outer boundary of the applicability of functional description.

7. I am taking working memory to be an active "workbench" or "blackboard" on which mental structures are constructed and manipulated, rather than just a passive store for rehearsal in the sense of Baddeley 1986, for instance. I gather my sense is not universally accepted. Readers should feel free to substitute their own favorite term for the functional capacity that builds sentence structures (other than "central executive," for if Fodor's (1983) notion of modularity is right about anything, it is that sentence perception is not a central executive function). See Jackendoff 2002a, sec. 7.3.

perception draws on anything akin to the learned store of words; the general presumption is that visual processing employs only more general (and unlearned) principles. However, some researchers (see e.g. Marr 1982; Ullman 1998; Cavanagh, Labianca, and Thornton 2001) have proposed that some higher-order visual processing is mediated in part by learned familiarity with certain kinds of objects and motions. Moreover, at least at the most central levels, much learning is necessary to establish cross-modal connections. For instance, everyone has learned an association between appearance and taste for hundreds if not thousands of kinds of food; we might consider this cross-modal knowledge a kind of "visual-to-taste lexicon" that helps interface between the two sorts of mental structures. Chapter 4 further proposes that there is a large "action lexicon," which encodes learned complex actions for purposes of both production and perception, and which links artifacts (e.g. doors, coffee-pots, faucets) with the appropriate actions performed using them.

A further kind of processing has to be mentioned: processes in which a structure on a particular level leads to construction of new structure(s) on the same level. The most prominent case is reasoning, which builds new conceptual structures from old. Such construction is governed by principles of inference (both logical and heuristic), which in this perspective are mappings from conceptual structures into further conceptual structures. But other mental processes might be treated as similar types of "within-level" construction, for instance mental rotation, which manipulates visual structures, and the computation of rhyme, which compares phonological structures.

Some general properties of the process of construction have emerged in research on language processing and also, I believe, in research on vision. First, construction is *incremental*: one does not need to flesh out a whole level before proceeding to the next. Rather, as soon as some structure is present in level L_1 that can be correlated with structure in the next level L_2, the interface linking L_1 and L_2 instantiates the correlated structure in L_2.

Second, construction is *promiscuous*: often the structure at a particular level is underdetermined by the process of construction up to that point in time. In general, the processor does not arbitrarily choose among the possibilities and then go on from there (as in the algorithmically conceived processing theories prevalent in the 1970s). Rather, it constructs *all* reasonable possibilities and runs them in parallel, eventually selecting a single most plausible or most stable structure as more constraints become available, and inhibiting the other structures. For instance, Swinney

(1979) and Tanenhaus, Leiman, and Seidenberg (1979) demonstrate that when a word in a sentence is first heard, all of its possible meanings are activated, whether contextually plausible or not; the otiose meanings are pared away over time as the word is integrated with the syntactic, semantic, and pragmatic context.[8]

Third, in order for such contextual effects to be possible, perception cannot be just bottom up. Rather, there has to be a degree of interaction in both directions. I find it useful to think of the process of construction as achieving a "resonance" among the linked structures, a state of global optimal stability within and among the structures in the complex. Occasionally among the promiscuous structures there are multiple stable states, in which case perception produces an ambiguous result such as the Necker cube in vision and a pun or other ambiguity in language.

Fourth, the processes of propagation through interfaces can in many cases be run in either direction. For instance, language perception is the process of beginning with a phonological structure and propagating structure to conceptual structure; and language production is the opposite, beginning with something to say (a conceptual structure) and propagating structure to phonology. The only part of the process that is unidirectional is the very periphery: one goes from audition to phonology and not the other way about, and one goes from phonology to motor control and not the reverse.

In vision, one is accustomed to thinking only in terms of perception, hence propagation of structure from sensory to central levels. But an instruction to imagine an elephant can provoke visual imagery. In such a case, the construction has to proceed from the interface(s) of the visual faculty with the most central levels of language structure. To the extent that noncentral levels of vision are involved in visual imagery (say if primary visual cortex is shown to be activated), propagation of activation has to be top-down. This means that except at the very periphery, visual processing too can be bidirectional.

1.5.3 Learning

A theory of mental function must be concerned not just with processing but also with learning. There are at least two different cases. An example of the first has already been mentioned: the taking in of information

---

8. The notion of promiscuity also shows up in Dennett's (1991) idea of the mind as constructing "multiple drafts," only one of which is selected to be the "narrative" in terms of which one understands and remembers one's current situation.

conveyed by an utterance. More generally, this is what goes under the term "one-time learning"—committing to long-term memory a structure that has been constructed in working memory. The formation of episodic memories would also fit under this rubric, if only we had a theory of the mental structures involved in perceiving and understanding "episodes."

The other type of learning, which might be called "slow(er) learning," involves the consolidation and generalization of material in long-term memory into schemas. For example, one approach to learning "rules of grammar" (e.g. Goldberg 1995, 2006; Culicover 1999; Jackendoff 2002a, secs. 6.9–6.10; Tomasello 2003; Culicover and Jackendoff 2005) conceives of them as abstractions from the structures of actual sentences the language learner has experienced. The formal difference between actual sentence structures and rules of grammar is that rules contain variables to be instantiated; the utterances one has experienced represent various instantiations of these variables. In other words, to learn a rule is to extract commonality among instances and replace the differences among the instances with a variable. In turn, rules of similar form can be further generalized, resulting in a stored schema with more and/or broader variables.

The result is a long-term memory that is more than a list of memories: it is structured in terms of "inheritance hierarchies," in which stored instances are at the bottom of the hierarchy and the most general schemas are at the top. This corresponds to a fairly broadly accepted sense of "semantic memory." However, in the "item-based" approach to language acquisition, inheritance hierarchies can be applied not only to semantic schemas such as *poodle → dog → animal → living thing*, but also to purely formal syntactic and phonological structures.

I take it that this type of "slow learning" is a process that takes place within long-term memory—as it were, a constant resifting of experience behind the scenes. But not much hangs on this.

1.6   A Caution, and What Modularity Means

It is common in cognitive neuroscience circles to speak of "information being broadcast through the brain." Here are some typical statements of this sort:

> ...the contents of awareness are to be understood as those information contents that are accessible to central systems, and brought to bear in a widespread way in the control of behavior. (Chalmers 1997, 22)

...conscious contents become 'globally available' to many unconscious systems. The reader's consciousness of *this phrase*, for example, makes *this phrase* available to interpretive systems that analyze its syntax and meaning, its emotional and motivational import, and its implications for thought and action. (Baars 1997, 241)

...it seems reasonable to hypothesize that awareness of a particular element of perceptual information must entail...access to that information by most of the rest of the mind/brain. (Kanwisher 2001, 105)

...dynamic mobilization makes [information available within a modular process] directly available in its original format to all other workspace processes. (Dehaene and Naccache 2001, 15)

On the view of mental structure and function being advocated here, this notion cannot be sustained. A phonological structure, for example, is intelligible "in its original format" only to the part of the mind/brain that processes phonological structure. If that part of the mind "broadcast" its contents to, say, a visual processor, it would be less than useless. And the same is true for any level of structure.

There is however a more restricted sense in which information is "broadcast." To the extent that a level of structure has interfaces to other levels, the interfaces can propagate activation to the related levels—but in the levels' own proprietary formats. So, for instance, phonological structure has an interface with syntax, so the presence of a phonological structure in working memory leads fairly automatically to the construction of a correlated syntactic structure. In turn, the syntactic structure interfaces with conceptual structure, so a linked triple of structures emerges over time. If the conceptual structure turns out to be an instruction to form a mental image, understanding the sentence leads to the propagation of structure into the visual system as well—in visual, not phonological format. But getting to the visual format requires passing through all the intervening interfaces.[9]

This is the sense in which I want to understand the notion of modularity. Each level of structure has its own proprietary format, incompatible with all the others. Thus it is "domain-specific" in the sense of Fodor 1983. Its interfaces with other levels of structure are what prevent it from

---

9. An exception: There is a specialized interface between phonology and vision that is responsible for *reading*. Thus a phonological structure *[elephant]* in working memory might lead to the construction of a visual image of the corresponding written form. The image of an elephant, however, would have to come by the normal route.

being functionally isolated in the mind. This leads to a relativized version of Fodor's notion of "informational encapsulation": a level L_1 is encapsulated from another level L_2 to the degree that distinctions in L_1 do not have direct correlates in L_2. For instance, phonological structure is less encapsulated from syntax than it is from spatial structure, in that phonological linear order and constituency correspond fairly closely with syntactic linear order and consituency, whereas the relation between phonology and spatial structure is far less direct, mediated by several intervening interfaces. (See Jackendoff 2002a, sec. 7.5, for more detailed comparison of this "structure-based" modularity with Fodorian modularity.)

I emphasize that all of this discussion of mental structure, outside of the language faculty, is strictly programmatic. Some of the later chapters of this book are in part an attempt to demonstrate the utility of this approach in studying other cognitive phenomena.