Fodor and Pylyshyn’s (1988, henceforth F&P) seminal work marked an important step forward in the debate about what kind of architecture is required for human cognition. By focusing on core concepts of “systematicity,” “symbolic,” and “compositionality,” F&P suggested new ways to consider the competing claims about connectionist versus symbolic systems in an explicit and precise way. Whether or not one agrees with their definitions (and there has been considerable debate both about the correctness as well as clarity and coherence of these definitions), the narrowing of the debate has led to stimulating discussion.

Absent such discussion, for example, terms such as ‘systematic’ may be taken to refer to significantly different things. The equation \( y(t) = f(x, y(t-1)) \), for example, describes a dynamical system that is, in general terms, systematic. The problem is that such equations can be used to characterize both connectionist and (with some stretching) symbolic systems. F&P usefully narrow the definition of systematic in a way that allows us to ask important questions: Are connectionist models capable of systematicity? Are only symbolic models systematic? Is human cognition itself systematic? The main point of F&P, of course, is that a cognitive architecture must be capable of systematicity in order to explain human cognition. A related formulation, which captures much of what F&P argued for, is that human cognition is algebraic, where that algebra requires symbols that are abstract and context free (see Marcus 2001).

Although F&P’s definitions characterize both representations as well as rules, much of the ensuing debate has tended to focus on their claims regarding systematicity of rules. Representations, and in particular, lexical representations, have not figured prominently in the debate. This emphasis on rules is not surprising, given the zeitgeist of the time in which the
paper was written. In that period, rules were seen as the major source of linguistic generativity. Because the lexicon was understood as involving a relatively stable inventory of entities with fixed meanings, pronunciations, and so on, the contribution of the lexicon to productive language use was felt to not be particularly interesting.

In intervening years, however, the lexicon has come into its own. Lexical knowledge is now acknowledged to be quite rich. Given proposals that the lexicon includes abstract patterns (e.g., Goldberg 2003), it is now seen as a source of linguistic productivity. One consequence of this is that boundary between the lexicon and rule systems has become somewhat blurred.

In this chapter, I consider ways in which systematicity, as F&P have defined it, might apply to the lexicon. Their treatment of the lexicon is light, which requires me to make some assumptions about how systematicity might apply to the lexicon. Whether or not these assumptions are reasonable is for the reader (and, presumably, F&P) to judge.

A number of chapters in this book challenge the claims made by F&P with regard to rule systems. I will argue that similar problems arise in considering the lexicon. So I begin by discussing empirical data that are problematic for the view of the lexicon as an enumerative database of the sort that would be required by an F&P approach. That is, the option of significantly enlarging the format of lexical entries to accommodate these data—the obvious move—not only leads to an unwieldy combinatoric explosion of information, but more seriously compromises the theoretical assumptions that motivated placing some, but not all, information in the lexicon. I then go on to argue that a very different mechanism than the lexicon is required in order to capture the richness of lexical knowledge. To lend some concreteness to this conclusion, I conclude with a preliminary model that I believe possesses some of the characteristics that are desirable in such a mechanism. This model eliminates the lexicon, qua lexicon, while providing an alternative way for lexical knowledge to play a role in language comprehension.

2 The Lexicon as Dictionary

The metaphor of the mental lexicon as a dictionary is pervasive and compelling. However, metaphors bring a lot of baggage with them, sometimes hidden from view. If the lexicon is to do real work for us, we need to go beyond metaphor and specify what the properties of the lexicon are.
There is considerable and important divergence on this issue, although commonalities exist across many theories. These are well captured in Jackendoff’s (2002) description of the lexicon:

For a first approximation, the lexicon is the store of words in long-term memory from which the grammar constructs phrases and sentences. (130)

[A lexical entry] lists a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics. (131)

A number of important questions immediately arise: (1) Just how small is “small”? (2) Why is there a limit on the quantity of this information? (3) What tests do we use to determine what information is included in the lexical representation, and what information resides elsewhere? There is no agreement on the answers to these questions. Indeed, these questions are rarely posed in an explicit form. However, there seems to be some implicit consensus that the limit on information is driven by notions of parsimony, and that only core information is included in lexical entries. “Core” may be defined operationally as information that is reliable and stable across context and is minimally sufficient to distinguish otherwise similar lexical entries.

Some awkwardness arises when dealing with polysemy. Many lexical items have multiple senses. In some cases, these senses share a common underlying meaning. The verb admit, for instance, can mean “to let in” as well as “to acknowledge.” Because these meanings are highly associated with different syntactic frames (the first prefers a direct object, the second a sentential complement), but because there is overlap in their meaning, they might appear as distinct senses in the same lexical entry. The very different meanings of the verb “bank,” on the other hand, suggest separate lexical entries that share the same phonological form. However, there are also many cases in which the distinctions in meaning may be quite subtle (cf. The journalist checked the facts vs. The mechanic checked the brakes; see Hare, Elman, and McRae 2000, for a fuller discussion of such facts and their effects on processing). The different uses of this same verb are associated with different preferences as to fillers of the patient role, but the senses are quite similar. Thus, although it is easy to come up with clear examples of clear polysemy versus homophony, there are many cases for which the vague notion of core provides no clear guide as to when two lexemes are polysemes or homophones.

In the psycholinguistics literature, an important additional processing assumption has emerged, often implicit but nonetheless important because
it suggests a test for what information might be included or excluded from a lexical representation: access to lexical information is faster and precedes the access of suprarexical (e.g., syntactic, pragmatic) information. For example, it has been argued that

lexical access is an autonomous subsystem of the sentence comprehension routine in which all meanings of a word are momentarily accessed, regardless of the factors of contextual bias or bias associated with frequency of use. (Onifer and Swinney 1981, 225)

Under this hypothesis, disambiguation occurs subsequent to the initial retrieval of all meanings, and extrarexical information is allowed to determine the contextually appropriate meaning at a later stage in processing. Despite the fact that subsequent work has challenged this claim (e.g., Van Petten and Kutas 1987, among many others), it remains a common assumption for many in the field that lexical information is privileged by rapid access. An important corollary of this is that early access to information is a good test of whether or not that information is contained in the lexicon. Factors that affect processing later are assumed to be extrarexical.

A final important assumption is that the lexicon is the bottom line as far as meaning is concerned. This is consistent with F&P’s definition of compositionality of meaning as being the molecular composition of atomic meanings that are indexed in the lexicon. Although the mapping between concepts and lexical semantics is not one to one, the relationship is in practice fairly tight.

3 Problems for This View of the Lexicon

I have already noted some challenges that are problematic for the lexicon as dictionary. These include the lack of specific tests for deciding what information is contained in lexical entries, the vagueness of the definition of “core,” and the gradations in meanings that often make it difficult to distinguish polysemy from homophony. I turn now to a broader set of facts that stretches the notion of what information should be contained in the lexicon, assuming that processing facts can be used to infer lexical versus extrarexical information. Because this information is often lexically idiosyncratic, it creates tension between information that is general and abstract and information that is lexically specific. I then move to a second set of facts that creates more significant challenges for the traditional lexicon (see Elman 2009 for a more extensive presentation), and which leads to my final proposal for an alternative way to encode lexical knowledge.
Although many theories of the lexicon begin with nouns, I will focus on verbs. Verbs are critical for binding other elements in a sentence together, and most of us have the intuition that, absent verbs, language is little more than a primitive indexical communication system.

### 3.1 Lexically Specific Knowledge

#### 3.1.1 The Relationship between Meaning and Complement Structure Preferences

One workhorse of the psycholinguistic literature has been the study of how comprehenders process sentences that contain a temporary ambiguity. In cases where competing theories predict how comprehenders will resolve this ambiguity, the comprehenders’ response when the ambiguity is first encountered, or subsequently disambiguated, provide useful clues as to what information was available and what processing strategy was used.

A much-studied structural ambiguity is that which arises at the postverbal noun phrase (NP) in sentences such as *The boy heard the story was interesting.* At the point where *the story* is encountered, it could either be the direct object (DO) of *heard*, or it could be the subject noun of a sentential complement (SC, as it ends up being in this example). An early influential theory of syntactic processing (the “two-stage model”; Frazier and Rayner 1982) predicts that the DO interpretation will be favored initially. Early data supported this claim. An alternative possibility (following constraint-based approaches) is that that result might arise for other reasons: (1) the relative frequency that a given verb occurs with either a DO or SC might favor the DO bias in this case (Garnsey et al. 1997; Holmes 1987; Mitchell and Holmes 1985); (2) the relative frequency that a given verb takes an SC with or without the disambiguating but optional complementizer *that* could lead to the DO bias (Trueswell, Tanenhaus, and Kello 1993); and (3) the plausibility of the postverbal NP as a DO for that particular verb (Garnsey et al. 1997; Pickering and Traxler 1998; Schmauder and Egan 1998).

The first of these factors—the statistical likelihood that a verb appears with either a DO or SC structure—has been particularly perplexing. The prediction is that if comprehenders are sensitive to the usage statistics of different verbs, then, when confronted with a DO/SC ambiguity, comprehenders will prefer the interpretation that is consistent with that verb’s bias. Some studies report either late or no effects of verb bias (e.g., Mitchell 1987; Ferreira and Henderson 1990). More recent studies, on the other hand, have shown that verb bias does affect comprehenders’ interpretation
of such temporarily ambiguous sequences (Garnsey et al. 1997; Trueswell et al. 1993; but see Kennison 1999). Whether or not such information is used at early stages of processing is important not only because of its processing implications but because, if it is, this implies that the detailed statistical patterns of subcategorization usage will need to be part of a verb’s lexical representation.

One possible explanation for the discrepant experimental data is that many of the verbs that show such DO/SC alternations have multiple senses, and these senses may have different subcategorization preferences (Roland and Jurafsky 1998, 2002). This raises the possibility that a comprehender might disambiguate the same temporarily ambiguous sentence fragment in different ways, depending on the inferred meaning of the verb. That meaning might in turn be implied by the context that precedes the sentence. A context that primes the sense of the verb that more frequently occurs with DOs should generate a different expectation than a context that primes a sense that has an SC bias.

Hare, McRae, and Elman (2003, 2004) tested this possibility. Several large text corpora were analyzed to establish the statistical patterns of usage that were associated with verbs (DO vs. SC) and in which different preferences were found for different verb senses. The corpus analyses were used to construct pairs of two sentence stories. In each pair, the second target sentence contained the same verb in a sequence that was temporarily (up to the postverbal NP) ambiguous between a DO or SC reading. The first sentence provided a meaning-biasing context. In one case, the context suggested a meaning for the verb in the target sentence that was highly correlated with a DO structure. In the other case, the context primed another meaning of the verb that occurred more frequently with an SC structure. Both target sentences were in fact identical until nearly the end. Thus, sometimes the ambiguity was resolved in a way that did not match participants’ predicted expectations. The data (reviewed in more detail in Hare et al. 2009) suggest that comprehenders’ expectancies regarding the subcategorization frame in which a verb occurs is indeed sensitive to statistical patterns of usage that are associated not with the verb in general, but with the sense-specific usage of the verb. A computational model of these effects is described by Elman, Hare, and McRae (2005).

A similar demonstration of the use of meaning to predict structure is reported by Hare, Elman, Tabaczynski, and McRae (2009). That study examined expectancies that arise during incremental processing of sentences that involve verbs such as collect, which can occur in either a transitive construction (e.g., The children collected dead leaves, in which the verb
has a causative meaning) or an intransitive construction (e.g., *The rainwater collected in the damp playground*, in which the verb is inchoative). Here again, at the point where the syntactic frame is ambiguous (at the verb, *The children collected*... or *The dead leaves collected*...), comprehenders appeared to expect the construction that was appropriate given the likely meaning of the verb (causative vs. inchoative). In this case, the meaning was biased by having subjects that were either good causal agents (e.g., *children* in the first example above) or good themes (*rainwater* in the second example).

These experiments suggest that the lexical representation of verbs must include not just information regarding the verb’s overall structural usage patterns, but that this information regarding the syntactic structures associated with a verb is sense-specific, and a comprehender’s structural expectations are modulated by the meaning of the verb that is inferred from the context. This implies a richer lexical representation for verbs than might have been assumed, though this can be easily accommodated within the traditional lexicon.

### 3.1.2 Verb-Specific Thematic Role Filler Preferences

Another well-studied ambiguity arises with verbs such as *arrest*. These are verbs that can occur in both the active voice (as in *The man arrested the burglar*) and in the passive (as in *The man was arrested by the policeman*). The potential for ambiguity arises because relative clauses in English (*The man who was arrested*...) may occur in a reduced form in which *who was* is omitted. This gives rise to *The man arrested*..., which is ambiguous at the verb. Until the remainder of the sentence is provided, it is temporarily unclear whether the verb is in the active voice (and the sentence might continue as in the first example) or whether this is the start of a reduced relative construction, in which the verb is in the passive (as in *The man arrested by the policeman was innocent*).

In an earlier study, Taraban and McClelland (1988) found that when participants read sentences involving ambiguous prepositional attachments, for example, *The janitor cleaned the storage area with the broom*... or *The janitor cleaned the storage area with the solvent*..., reading times were faster in sentences involving more typical fillers of the instrument role (in these examples, *broom* rather than *solvent*). McRae, Spivey-Knowlton, and Tanenhaus (1998) noted that in many cases, similar preferences appear to exist for verbs that can appear in either the active or passive voice. For many verbs, there are nominals that are better fillers of the agent role than the passive role, and vice versa.
This led McRae et al. (1998) to hypothesize that when confronted with a sentence fragment that is ambiguous between a main verb and reduced relative reading, comprehenders might be influenced by the initial subject NP and whether it is a more likely agent or patient. In the first case, this should encourage a main verb interpretation; in the latter case, a reduced relative should be favored. This is precisely what McRae et al. found to be the case. *The cop arrested*… promoted a main verb reading over a reduced relative interpretation, whereas *The criminal arrested*… increased the likelihood of the reduced relative reading. McRae et al. concluded that the thematic role specifications for verbs must go beyond simple categorical information, such as agent, patient, instrument, beneficiary, and so on. The experimental data suggest that the roles contain very detailed information about the preferred fillers of these roles, and that the preferences are verb specific. The preferences are expressed not only over the nominal fillers of roles, but their attributes as well. Thus, a shrewd, heartless gambler is a better agent of manipulate than a young, naive gambler; conversely, the latter is a better filler of the same verb’s patient role (McRae, Ferretti, and Amyote 1997).

This account of thematic roles resembles that of Dowty (1991) in that both accounts suggest that thematic roles have internal structure. But the McRae et al. (1997; McRae, Spivey-Knowlton, and Tanenhaus 1998) results further suggest a level of information that goes considerably beyond the limited set of proto-role features envisioned by Dowty. McRae et al. interpreted these role-filler preferences as reflecting comprehenders’ specific knowledge of the event structure associated with different verbs. This appeal to event structure, as we shall see below, will figure significantly in phenomena that are not as easily accommodated by the lexicon.

We have seen that verb-specific preferences for their thematic role fillers arise in the course of sentence processing. Might such preferences also be revealed in word-word priming? The question is important because this sort of priming has often been assumed to occur at the lexical level, that is, to reflect the ability of one word to activate another word, and thus to be a test of the context of a word’s lexical entry.

The answer is yes, such priming does occur. Ferretti, McRae, and Hatherell (2001) found that verbs primed nouns that were good fillers for their agent, patient, or instrument roles. The priming also goes in the opposite direction, such that when a comprehender encounters a noun, the noun serves as a cue for the event in which it typically participates, thereby priming verbs that describe that event activity (McRae et al. 2005).
The above results are among a much larger empirical literature that significantly extend the nature of the information that must be encoded in a verb’s lexical representation. In addition to sense-specific structural usage patterns, the verb’s lexical entry must also encode verb-specific information regarding the characteristics of the nominals that best fit that verb’s thematic roles. The lexical representation for verbs must include subentries about all the verb’s senses. For each sense, all possible subcategorization frames would be shown. For each verb-sense-subcategorization combination, additional information would be indicating the probability of each combination. Finally, similar information would be needed for every verb-sense-thematic role possibility. The experimental evidence indicates that in many cases, this latter information will be detailed, highly idiosyncratic of the verb, and represented at the featural level (e.g., Ferretti, McRae, and Hatherell 2001; McRae, Ferretti, and Amyote 1997).

3.2 Flies in the Ointment
These findings, among many others in recent years, expand the contents of the verb’s lexical representation. But even though these data suggest very detailed and often idiosyncratic lexical representations, they could still be accommodated by an enumerative data structure of the sort implemented by the lexicon. We now turn to additional phenomena that are problematic for the traditional view of the mental lexicon qua dictionary.

3.2.1 Aspect and Event Knowledge
As noted above, Ferretti et al. (2001) found that verbs were able to prime their preferred agents, patients, and instruments. However, no priming was found from verbs to the locations in which their associated actions take place. Why might this be? Ferretti, Kutas, and McRae (2007) noted that in that experiment the verb primes for locations were in the past tense (e.g., skated—arena), and possibly interpreted by participants as having perfective aspect. Because the perfective signals that the event has concluded, it is often used to provide background information prefatory to the time period under focus (as in Dorothy had skated for many years and was now looking forward to her retirement). Imperfective aspect, on the other hand, is used to describe events that are either habitual or ongoing; this is particularly true of the progressive. Ferretti et al. hypothesized that although a past perfect verb did not prime its associated location, the same verb in the progressive might do so because of the location’s greater salience to the unfolding event.
This prediction was borne out. The two-word prime had skated failed to yield significant priming for arena in a short SOA naming task, relative to an unrelated prime; but the two-word prime was skating did significantly facilitate naming. In an ERP version of the experiment, the typicality of the location was found to affect expectations. Sentences such as The diver was snorkeling in the ocean (typical location) elicited lower amplitude N400 responses at ocean, compared to The diver was snorkeling in the pond at pond. The N400 is interpreted as an index of semantic expectancy, and the fact that typicality of agent-verb-location combinations affected processing at the location indicates that this information must be available early in processing.

The ability of verbal aspect to manipulate sentence processing by changing the focus on an event description, with implications for processing, has been noted elsewhere (e.g., Kehler 2002; Kehler et al. 2008). The results in this case, however, present a specific challenge for how to represent verb argument preferences. Critically, the effect seems to occur on the same time scale as other information that affects verb argument expectations (this was demonstrated by Experiment 3 in Ferretti et al. 2007, in which ERP data indicated aspectual differences within 400 ms of the expected word’s presentation). This is a time frame that has often been seen as indicating that intralexical information is operant, and prior to adjustments that depend on extralexical (e.g., semantic, discourse, pragmatic) factors. But logically, it is difficult to see how one would encode the dynamic and context-specificity contingency on thematic role requirements that arises when aspect is manipulated. That is, although the patterns of ambiguity resolution described in earlier sections, along with parallel findings using priming (Ferretti, McRae, and Hatherell 2001; McRae et al. 2005), might be accommodated by enriching the information in the lexical representations of verbs, the very similar effects of aspect do not seem amenable to such an account.

Setting this important question aside for the moment (we return to it later), we might ask, If verb aspect can alter the expected arguments for a verb, what else might do so? The concept of event representation has emerged as a useful way to understand other results in which aspect plays a role (Kehler 2002; Kehler et al. 2008; Kertz, Kehler, and Elman 2006; Moens and Steedman 1988; Rohde, Kehler, and Elman 2006). If we consider the question from the perspective of event representation, viewing the verb as providing merely some of the cues (albeit very potent ones) that tap into event knowledge, then several other candidates suggest themselves.
3.2.2 Dynamic Alterations in Verb Argument Expectations

If we think in terms of verbs as cues and events as the knowledge they target, then it should be clear that although the verb is obviously a very powerful cue, and its aspect may alter the way the event is construed, there are other cues that change the nature of the event or activity associated with the verb. For example, the choice of agent of the verb may signal different activities. A sentence-initial noun phrase such as *The surgeon*... is enough to generate expectancies that constrain the range of likely events. In isolation, this cue is typically fairly weak and unreliable, but different agents may combine with the same verb to describe quite different events.

Consider the verb *cut*. Our expectations regarding what will be cut, given a sentence that begins *The surgeon cuts*... are quite different than for the fragment *The lumberjack cuts*... These differences in expectation clearly reflect our knowledge of the world. This is not remarkable. The critical questions are: What is the status of such knowledge, and where does it reside? No one doubts that a comprehender’s knowledge of how and what a surgeon cuts, versus what a lumberjack cuts, plays an important role in comprehension at some point.

The crucial issue, for the purposes of deciding what information is included in a lexical entry and what information arises from other knowledge sources, is when this knowledge enters into the unfolding process of comprehension. This is because, as pointed out above, timing has been an important adjudicator for models of processing and representation. If the knowledge is available very early—perhaps even immediately on encountering the relevant cues—then it is a candidate for being present in the lexical representation.

Bicknell, Elman, Hare, McRae, and Kutas (2010) hypothesized that if different agent-verb combinations imply different types of events, this might lead comprehenders to expect different patients for the different events. This prediction follows from a study by Kamide, Altmann, and Haywood (2003). Kamide et al. employed a paradigm in which participants’ eye movements toward various pictures were monitored as they heard sentences such as *The man will ride the motorbike* or *The girl will ride the carousel* (all combinations of agent and patient were crossed) while viewing a visual scene containing a man, a girl, a motorbike, a carousel, and candy. At the point when participants heard *The man will ride*..., Kamide et al. found that there were more looks toward the motorbike than to the carousel, and the converse was true for *The girl will ride*.... The Bicknell et al. study was designed to look specifically at agent-verb interactions.
to see whether such effects also occurred during self-paced reading, and if so, how early in processing.

A set of verbs such as cut, save, and check were first identified as potentially describing different events depending on the agent of the activity, and in which the event described by the agent-verb combination would entail different patients. These verbs were then placed in sentences in which the agent-verb combination was followed either by the congruent patient, as in The journalist checked the spelling of his latest report or in which the agent-verb was followed by an incongruent patient, as in The mechanic checked the spelling of his latest report (all agents of the same verb appeared with all patients, and a continuation sentence followed that increased the plausibility of the incongruent events). Participants read the sentences a word at a time, using a self-paced moving-window paradigm.

As predicted, reading times increased for sentences in which an agent-verb combination was followed by an incongruent (though plausible) patient. The slowdown occurred at one word following the patient, leaving open the possibility that the expectation reflected delayed use of world knowledge. Bicknell et al. therefore carried out a second experiment using the same materials, but recording ERPs as participants read the sentences. The rationale for this was that ERPs provide a more precise and sensitive index of processing than reading times. Of particular interest was the N400 component, since this provides a good measure of the degree to which a given word is expected and/or integrated into the prior context. As predicted, an elevated N400 was found for incongruent patients.

The fact that what patient is expected may vary as a function of specific particular agent-verb combinations is not in itself surprising. What is significant is that the effect occurs at the earliest possible moment, at the patient that immediately follows the verb. The timing of such effects has in the past often been taken as indicative of an effect’s source. A common assumption has been that immediate effects reflect lexical or “first-pass” processing, and later effects reflect the use of semantic or pragmatic information. In this study, the agent-verb combinations draw on comprehenders’ world knowledge. The immediacy of the effect would seem to require either that this information must be embedded in the lexicon or that world knowledge must be able to interact with lexical knowledge more quickly than has often typically been assumed.

Can other elements in a sentence affect the event type that is implied by the verb? Consider again the verb cut. The Oxford English Dictionary shows the transitive form of this verb as having a single sense. WordNet gives 41 senses. The difference is that WordNet’s senses more closely cor-
respond to what one might call event types, whereas the *OED* adheres to a more traditional notion of sense that is defined by an abstract core meaning that does not depend on context. Yet cutting activities in different contexts may involve quite different sets of agents, patients, instruments, and even locations. The instrument is likely to be a particularly potent constraint on the event type.

Matsuki, Chow, Hare, Elman, Scheepers, and McRae (2011) tested the possibility that the instrument used with a verb would cue different event schemas, leading to different expectations regarding the most likely patient. Using eye-tracking to monitor processing during reading, participants were presented with sentences such as *Susan used the scissors to cut the expensive paper that she needed for her project*, or *Susan used the saw to cut the expensive wood*.... Performance on these sentences was contrasted with that on the less expected *Susan used the scissors to cut the expensive wood*... or *Susan used the saw to cut the expensive paper*.... As in the Bicknell et al. study, materials were normalized to ensure that there were no direct lexical associations between instrument and patient. An additional priming study was carried out in which instruments and patients served as prime-target pairs; no significant priming was found between typical instruments and patients (e.g., scissors-paper) versus atypical instruments and patients (e.g., saw-paper; but priming did occur for a set of additional items that were included as a comparison set). As predicted, readers showed increased reading times for the atypical patient relative to the typical patient. In this study, the effect occurred right at the patient, demonstrating that the filler of the instrument role for a specific verb alters the restrictions on the filler of the patient role.

4 Lexical Knowledge without a Lexicon

4.1 Where Does Lexical Knowledge Reside?

The findings reviewed in section 3.1 strongly support the position that lexical knowledge is quite detailed, often idiosyncratic and verb specific, and brought to bear at the earliest possible stage in incremental sentence processing. The examples above focused on verbs and the need to encode restrictions (or preferences) over the various arguments with which they may occur. Taken alone, those results might be accommodated by simply providing greater detail in lexical entries in the mental lexicon, as standardly conceived.

Where things get tricky is when one also considers what seems to be the ability of dynamic factors to significantly modulate such expectations
(section 3.2). These include the verb’s grammatical aspect, the agent and instrument that are involved in the activity, and the overall discourse context. To be clear: that these factors play a role in sentence processing is not itself surprising. However, the common assumption has been that such dynamic factors lie outside the lexicon. This is, for example, essentially the position outlined by J. D. Fodor (1995): “We may assume that there is a syntactic processing module, which feeds into, but is not fed by, the semantic and pragmatic processing routines ... syntactic analysis is serial, with back-up and revision if the processor’s first hypothesis about the structure turns out later to have been wrong” (435).

More pithily, the data do not accord with the “syntax proposes, semantics disposes” hypothesis (Crain and Steedman 1985). Thus, what is significant about the findings above is that the influence of aspect, agent, instrument, and discourse all occur within the same time frame that has been used operationally to identify information that resides in the lexicon. This is important if we are to have some empirical basis for deciding what goes in the lexicon and what does not.

All of this places us in the uncomfortable position of having to make some difficult decisions.

One option would be to abandon any hope of finding any empirical basis for determining the contents of the mental lexicon. One might simply stipulate that some classes of information reside in the lexicon and others do not. This is not a desirable solution. Note that even within the domain of theoretical linguistics, considerable controversy has emerged regarding what sort of information belongs in the lexicon, with different theories taking different and often mutually incompatible positions (cf., among many other examples, Haiman 1980; Lakoff 1971; Weinreich 1962; Jackendoff 1983, 2002; Katz and Fodor 1963; Langacker 1987; Chomsky 1965; Levin and Hovav 2005; Fodor 2002). If we insist that the form of the mental lexicon has no consequences for processing, and exclude data of this type, then we have no behavioral way to evaluate different proposals. This essentially accepts that performance tells us little about competence (Chomsky 1965).

A second option would be to significantly enlarge the format of lexical entries so that they accommodate all the above information. This would be a logical conclusion to the trend that has appeared not only in the processing literature (e.g., in addition to the studies cited above, van Berkum et al. 2003; van Berkum et al. 2005; Kamide, Altmann, and Haywood 2003; Kamide, Scheepers, and Altmann 2003; Altmann and Kamide 2007) but also many recent linguistic theories (e.g., Bresnan 2006;
Fauconnier and Turner 2002; Goldberg 2003; Lakoff 1987; Langacker 1987; though many or perhaps all of these authors might not agree with such a conclusion). The lexicon has become increasingly rich and detailed in recent years. Why impose arbitrary limits on its contents?

One problem is that the combinatoric explosion this entails, especially given the unbounded nature of discourse contexts, may render the proposal infeasible. But it also presents us with a logical conundrum: if all this information resides in the lexicon, is there then any meaningful distinction between the lexicon and other linguistic modules?

The third option is the most radical: is it possible that lexical knowledge of the sort discussed here might be instantiated in a very different way than through an enumerative dictionary?

4.2 An Alternative to the Mental Lexicon as Dictionary

The common factor in the studies described above was the ability of sentential elements to interact in real time to produce an incremental interpretation that guided expectancies about upcoming elements. These can be thought of as very powerful context effects that modulate the meaning that words have.

But suppose that one views words not as elements in a data structure that must be retrieved from memory, but rather as stimuli that alter mental states (which arise from processing prior words) in lawful ways. In this view, words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not have meaning; rather, they are cues to meaning (Elman 2009; Rumelhart 1979).

This scheme of things can be captured by a model that instantiates a dynamical system. The system receives inputs (words, in this case) over time. The words perturb the internal state of the system (we can call it the “mental state”) as they are processed, with each new word altering the mental state in some way.

Over the years, a number of connectionist models have been developed that illustrate ways in which context can influence processing in complicated but significant ways (e.g., among many others, McClelland and Rumelhart 1981; McRae, Spivey-Knowlton, and Tanenhaus 1998; Rumelhart et al. 1988; Taraban and McClelland 1988). There is also a rich literature in the use of dynamical systems to model cognitive phenomena (e.g., Smith and Thelen 1993; Spencer and Schöner 2003; Tabor and Tanenhaus 2001; Thelen and Smith 1994).

A particularly fruitful architecture has been one that involves recurrence between processing units (e.g., Botvinick and Plaut 2004; Elman 1990;
Rogers et al. 2004; St. John and McClelland 1990). In recurrent networks, information flow is multidirectional, and feedback loops allow the current state of the system to be affected by prior states. Learning in such systems can be thought of as encoding the grammar over sequences. The grammar constrains the lawful effects that inputs have on moving the system through the network’s “mental state space.” If those inputs are words, then what we think of as lexical knowledge is the knowledge encoded in the connections between processing units that allows each word to have the appropriate effect on processing. We have not removed the need for lexical knowledge, but rather moved it from an enumerative and declarative database (the lexicon) into the elements of the system (the weights) that are responsible for processing. It should be noted that recurrent networks are not finite state automata, but have computational properties similar—but not identical—to stack-based automata (Boden and Wiles 2000; Boden and Blair 2003; Peters and Ritchie 1973; Rodriguez, Wiles, and Elman 1999; Rodriguez and Elman 1999; Rodriguez 2001; Siegelmann and Sontag 1995).

Elman (2009) presented a simple model that demonstrates some of the properties required to capture the kinds of lexical effects noted here. A somewhat fuller model, under development in collaboration with Ken McRae and Mary Hare, is shown in figure 5.1.

The model is inspired by and incorporates elements of a number of important prior models that have related properties. These include models of language processing (McClelland, St. John, and Taraban 1989; St. John and McClelland 1990; St. John 1992), schemas and sequential thought processes (Rumelhart et al. 1988), semantic cognition (Rogers et al. 2004), and action planning (Botvinick and Plaut 2004). One way of thinking of this model is as an attempt to take the important insights regarding schemas, scripts, frames, and stories (Abelson 1981; Minsky 1974; Norman and Rumelhart 1981; Schank and Abelson 1977) and instantiate those insights in a computational architecture that allows for richer processing than was possible using the earlier tools from the AI toolbox.

The goal of the model is to learn the contingent relationships between activities and participants that are involved in events that unfold over time. The model has a view of the world (i.e., inputs) that allows it to identify the relevant entities (including, in this simplified version, agents, patients, instruments, and locations) and actions that participate in momentary activities (e.g., *John enters the restaurant*). These activities are connected in sequence to form events (e.g., *John enters the restaurant; He sits down at a table; He orders food; He cuts the food with a knife; He eats the food;…; He leaves*). At any given point in time, the task given to the model
An activity consists of a collection of entities, actions, and context, all presented simultaneously as input from the world. The “current activity” portion of the network learns the patterns of co-occurrence that are typical of given activities (e.g., *Jill cuts her steak with a knife and fork in the restaurant*). The recurrent connections between the entity-action-context units and the hidden units allow the network to complete input patterns in which not all information is explicitly specified. Thus, *Jill cuts her steak* would lead to the activation of *knife* and *restaurant*. An event consists of a sequence of activities, presented in temporal order and duration appropriate to the event. The “predicted next activity” portion of the network learns the temporal relations between activities that are part of an event. This part of the network also learns to do pattern completion in the temporal domain. Thus, if the current activity is *Jill enters the restaurant*, the predicted next activity layer would anticipate subsequent activities, and would activate, in sequence, *Jill orders a steak*, then *Jill cuts her steak with a knife and fork*, then *Jill eats her steak*, and so on.
is to learn the mutual constraints between co-occurring entities and actions (this is accomplished by the model’s reproducing on its output layer the details of the current activity), and also to generate expectancies regarding subsequent activities that together compose this event (accomplished by the model’s predictions of future activities).

During learning, the model may see only a subset of possible entities and actions that make up an activity, or a subset of the sequence of activities that make up an event. Over time, however, accumulated experience allows the network to learn about the overarching generalizations that exist between and across events, and to fill in missing information as required. This leads to the following characteristics:

1. Pattern completion within activities and within events
2. Typicality and prototype effects
3. Soft and graded constraints on roles, participants, activities, locations, and so on
4. Ability to flexibly combine and merge novel combinations of events
5. Ability to support inferences—under the right conditions
6. Ability to capture effects of perspective on event representation

The details of the model and simulations of its behavior are described elsewhere (Elman, McRae, and Hare, in preparation), and given the scope of this chapter, I present two simulations to illustrate these properties. Central to all of these is the fact that the model implements a constraint-satisfaction network. The constraints operate both at a given point in time, reflecting patterns of co-occurrence of actions and activity participants, and across time, reflecting the succession of activities that arise over time as an event unfolds.

In figure 5.2, we see the activation of various entities and actions that result from the model’s being presented with John does something with his food. This activity occurs in a sequence shortly after John has entered a restaurant. The activation levels are indicated by the height of curves on the ordinate, and their change over time is indicated along the abscissa. The model infers that food is involved throughout the activity, that John begins by cutting the food (so cut and knife are active), and that, as time goes along, the cutting action diminishes and the eating action increases. The model has thus not only filled in the implied elements but has captured the ordered temporal relationships between activities that make up the overall event.

These properties provide a straightforward account of the findings reported in Metusalem et al. (2012), in which words that were unexpected
but consistent with the larger event being described elicited smaller amplitude N400s than words that were both unexpected and unrelated to the event. In figure 5.3, we see the time-varying activation of various elements as an event is being described. At the end, the activation of medal, which is event relevant even though it is locally unexpected given the linguistic context, is higher than bleach, which is both unexpected and event unrelated.

In other simulations, we find that the model is able to use its knowledge of events, with the resulting activation of event-relevant participants that may not be named, to make inferences that affect subsequent ambiguous statements. For example, compare the following two sequences:

1.  *John cut wood in the forest.*
   *Suddenly, he cut himself by mistake.*
   *What happened to John?*

2.  *John cut food in the restaurant.*
   *Suddenly, he cut himself by mistake.*
   *What happened to John?*
The crowd was seated in the stands, watching the skater basking in glory. The skater went to the podium. The crowd applauded. The skater smiled as the national anthem was played.

**Figure 5.3**

Time-varying activation of three entities (medal, podium, bleach) as the network processes four activities associated with the event “getting a medal at the Olympics.” In the final activity, none of the three entities is mentioned. However, podium is reactivated (following a drop during the crowd applauding activity), because that location has previously been associated with the awarding of medals. Medal is also activated, even though it is contextually unexpected given the linguistic context, because it is consistent with the event. In contrast, bleach, which is equally unexpected in this linguistic context, is completely inactive because it is inappropriate to the event. These patterns replicate the findings of Metusalem et al. (2012).

In the context of cutting wood in the forest, the model has learned that axes are typically used. It also knows, independently (from other event knowledge), that cutting oneself with an axe can be fatal. Cutting food in a restaurant, on the other hand, involves knives; and cutting oneself with a knife is less lethal. When confronted with the final (identical) query, the model activates *John died* in the first scenario and *John bled* in the second case. The bridging inference is made possible by the model’s ability not only to fill in the missing pieces (cutting wood in the forest probably involves an axe, etc.), but then to use the inferred but unmentioned information to resolve a subsequent ambiguity (*What happened?*)
5 Discussion

Although the possibility of lexical knowledge without a lexicon might seem odd, the core ideas that motivate this proposal are not new. Many elements appear elsewhere in the literature. These include the following.

(1) The meaning of a word is rooted in our knowledge of both the material and the social world. The material world includes the world around us as we experience it (i.e., it is embodied), possibly indirectly. The social world includes cultural habits and artifacts; in many cases, these habits and artifacts have significance only by agreement (i.e., they are conventionalized). Similar points have been made by many others, notably Wittgenstein (1966), Hutchins (1994) and Fauconnier (1997), and Fauconnier and Turner (2002).

(2) Context is always with us. The meaning of a word is never “out of context,” although we might not always know what the context is (particularly if we fail to provide one). This point has been made by many, including Kintsch (1988), Langacker (1987), McClelland et al. (1989), and van Berkum and colleagues (van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort 2005; van Berkum, Zwitserlood, Hagoort, and Brown 2003). This insight is also found in computational models of meaning that emphasize multiple co-occurrence constraints between words in order to represent them as points in a high dimensional space, such as LSA (Latent Semantic Analysis) (Landauer and Dumais 1997), HAL (Burgess and Lund 1997), or probabilistic models (Griffiths and Steyvers 2004). The dynamical approach here also emphasizes the time course of processing that results from the incremental nature of language input.

(3) Events play a major role in organizing our experience. Event knowledge is used to drive inference and access memory, and it affects the categories we construct. An event may be defined as a set of participants, activities, and outcomes that are bound together by causal interrelatedness. An extensive literature argues for this, aside from the studies described here, including work by Minsky (1974), Schank and Abelson (1977), and Zacks and Tversky (2001); see also Shipley and Zacks’s (2008) book for a comprehensive collection on the role of event knowledge in perception, action, and cognition.

(4) Dynamical systems provide a powerful framework for understanding biologically based behavior. The nonlinear and continuous valued nature of dynamical systems allows them to respond in a graded manner under
some circumstances, while in other cases their responses may seem more binary. Dynamical analyses figure prominently in the recent literature in cognitive science, including work by Smith and Thelen (1993, 2003; Thelen and Smith 1994), Spencer and Schöner (2003), Spivey (2007; Spivey and Dale 2004), and Tabor (2004; Tabor and Tanenhaus 2001; Tabor et al. 1997).

The model presented here is a first step in trying to offer an alternative mechanism for representing lexical knowledge. Rather than assuming that knowledge is enumerated in a database, the model moves the knowledge into the dynamics that result from a word’s effect on processing. These dynamics are lawful and predictable. They might even be called systematic, although probably not under the definition offered by F&P. Under the model’s version of systematicity, such things as context effects, graded polysemy, and lexically idiosyncratic behavior are not seen as exceptions that require ad hoc explanation. Rather, these properties are predicted by the model. At the same time, the model is able to generalize and abstract across commonalities of behavior. This is “having your cake and eating it too.”

The model is incomplete in important ways. It is disembodied and so lacks the conceptual knowledge about events that comes from direct experience (of course, this deficiency is equally true of F&P symbolic models). The model assumes a prior identification of some entities as agents, and others as patients, instruments, and so on. Proposals have in fact been put forward about how this might be done (e.g., Chang, Dell, and Bock 2006; Chang 2012; Gao, Newman, and Scholl 2009; Gao and Scholl 2011). An interesting future direction will be to extend the model in similar ways.

How does this view affect the way we do business (or at least, study words)? Although I have argued that many of the behavioral phenomena described above are not easily incorporated into a mental lexicon as traditionally conceived, it is possible that solutions exist. A parallel architecture of the sort described by Jackendoff (2002), for example, if it permitted direct and immediate interactions among the syntactic, semantic, and pragmatic components of the grammar, might be able to account for the data described earlier. The important question would then be how to devise tests to distinguish between these proposals. This remains an open question for the moment, at least until such counterproposals are advanced.

However, theories are also evaluated for their ability to offer new ways of thinking about old problems, and to provoke new questions that would not be otherwise asked. Let me suggest two positive consequences to the
sort of words-as-cues dynamical model I am outlining (see also Elman 2009, for a fuller discussion).

The assumption that only certain information goes in the lexicon, and that the lexicon and other knowledge sources respect modular boundaries with limited and late occurring interactions, drives a research program that discourages looking for evidence of richer and more immediate interactions. For example, that selectional restrictions might be dynamic and context sensitive is fundamentally not an option within the Katz and Fodor (1963) framework. The words-as-cues approach, in contrast, suggests that such interdependencies should be expected. Indeed, there should be many such interactions among lexical knowledge, context, and nonlinguistic factors, and these might occur early in processing. Many researchers in the field have already come to this point of view. It is a conclusion that, despite considerable empirical evidence, has taken longer to arrive than it might have, given a different theoretical perspective.

A second consequence of this perspective is that it encourages a more unified view of phenomena that are often treated (de facto, if not in principle) as unrelated. Syntactic ambiguity resolution, lexical ambiguity resolution, pronoun interpretation, text inference, and semantic memory (to choose but a small subset of domains) are studied by communities that do not always communicate well, and researchers in these areas are not always aware of findings from other areas. Yet these domains have considerable potential for informing each other. That is because, although they ultimately draw on a common conceptual knowledge base, that knowledge base can be accessed in different ways, and this in turn affects what is accessed. Consider how our knowledge of events might be tapped in a priming paradigm, compared with a sentence-processing paradigm. Because prime-target pairs are typically presented with no discourse context, one might expect that a transitive verb prime might evoke a situation in which the fillers of both its agent and patient roles are equally salient. Thus, arresting should prime cop (typical arrestor) and also crook (typical arrestee). Indeed, this is what happens (Ferretti, McRae, and Hatherell 2001). Yet this same study also demonstrated that when verb primes were embedded in sentence fragments, the priming of good agents or patients was contingent on the syntactic frame within which the verb occurred. Primes of the form She arrested the... facilitated naming of crook, but not cop. Conversely, the prime She was arrested by the... facilitated naming of cop rather than crook.

These two results demonstrate that although words in isolation can serve as cues to event knowledge, they are only one such cue. The grammatical construction within which they occur provides independent
evidence regarding the roles played by different event participants (Goldberg 2003). And, of course, the discourse context may provide further constraints on how an event is construed. Thus, as Race et al. (2008) found, although shoppers might typically save money and lifeguards save children, in the context of a disaster, both agents will be expected to save children.

Eliminating the lexicon is indeed radical surgery, and it is an operation that at this point many will not agree to. At the very least, however, I hope that by demonstrating that lexical knowledge without a lexicon is possible, others will be encouraged to seek out additional evidence for ways in which the many things that language users know are brought to bear on the way language is processed.

Acknowledgments

This work was supported by NIH grants HD053136 and MH60517 to Jeff Elman, Ken McRae, and Mary Hare; NSERC grant OGP0155704 to KM; NIH Training Grant T32-DC000041 to the Center for Research in Language (UCSD); and by funding from the Kavli Institute of Brain and Mind (UCSD). Much of the experimental work reported here and the insights regarding the importance of event knowledge in sentence processing are the fruit of a long-time productive collaboration with Mary Hare and Ken McRae. I am grateful to them for many stimulating discussions and, above all, for their friendship. I am grateful to Jay McClelland for the many conversations we have had over three decades. His ideas and perspectives on language, cognition, and computation have influenced my thinking in many ways. Finally, a tremendous debt is owed to Dave Rumelhart, whose (1979) suggestion that words do not have meaning but rather are cues to meaning inspired the proposal outlined here.

References


