Running Head: ON WORDS AND DINOSAUR BONES

On the meaning of words and dinosaur bones:

Lexical knowledge without a lexicon

Jeffrey L. Elman

University of California, San Diego

La Jolla, CA

Correspondence concerning this article should be addressed to:

Jeffrey L. Elman

Department of Cognitive Science, 0526

University of California, San Diego

La Jolla, CA 92093-0526

Email: jelman@ucsd.edu

Phone: (858) 534-1147

Fax: (858) 822-5097

Abstract

Although for many years a sharp distinction has been made in language research between rules and words—with primary interest on rules—this distinction is now blurred in many theories. If anything, the focus of attention has shifted in recent years in favor of words. Results from many different areas of language research suggest that the lexicon is representationally rich, that it is the source of much productive behavior, and that lexically-specific information plays a critical and early role in the interpretation of grammatical structure. But how much information can or should be placed in the lexicon? This is the question I address here. I review a set of studies whose results indicate that event knowledge plays a significant role in early stages of sentence processing and structural analysis. This poses a conundrum for traditional views of the lexicon. Either the lexicon must be expanded to include factors that do not plausibly seem to belong there; or else virtually all information about word meaning is removed, leaving the lexicon impoverished. I suggest a third alternative, which provides a way to account for lexical knowledge without a mental lexicon.

For a first approximation, the lexicon is the store of words in long-term memory from which the grammar constructs phrases and sentences. [A lexical entry is] lists a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics.

Ray Jackendoff

My approach suggests that comprehension, like perception, should be likened to Hebb's (1949) paleontologist, who uses his beliefs and knowledge about dinosaurs in conjunction with the clues provided by the bone fragments available to construct a full-fledged model of the original. In this case the words spoken and the actions taken by the speaker are likened to the clues of the paleontologist, and the dinosaur, to the meaning conveyed through these clues.

David Rumelhart

1. Introduction

I begin with a warning to the reader. I propose to do away with one of the objects most cherished by language researchers: The mental lexicon. I do not call into question the existence of words, nor the many things language users know about them. Rather, I suggest the possibility of lexical knowledge without a lexicon.

Words have made a comeback in the past few decades. For many years, linguistic theories focused primarily on rules. The apparently idiosyncratic character of words (in the sense that the mapping between meaning and form was arbitrary and varied randomly across languages) made them relatively uninteresting for many language researchers. Rules were where the action was.

Over the years, however, many linguists have come to see words not simply as flesh that gives life to grammatical structures, but as bones that are themselves grammatical rich entities. This sea change has accompanied the rise of usage-based theories of language (e.g., Langacker, 1987; Tomasello, 2003), which emphasize the context-sensitivity of word use. In some theories,

the distinction between rule and word is blurred, with both seen as objects that implement form-mapping relationships (Goldberg, 2003; Jackendoff, 2007). Within developmental psychology, words have always been of interest (after all, In the beginning, there was the word…) but more recent theories suggest that words may themselves be the foundational elements from which early grammar arises epiphenomenally (Bates & Goodman, 1997; Tomasello, 2000). In the field of psycholinguistics, an explosion of findings indicate that interpretation of a sentence's grammatical structure interacts with the comprehender's detailed knowledge of properties of the specific words involved. Furthermore, these interactions occur at early stages of processing (Altmann, 1998; MacDonald, 1997; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). All of this suggests that lexical representations are quite rich and detailed, that their content arises from habits of usage, and that this wealth of lexically-specific information affects the interpretation of higher-level grammatical structure at very early stages of processing.

Representation is a matter not only of content, however, but also of form. As the representational content of words increases, one might wonder whether this content at some point exceeds the capacity of the assumed form of lexical items to contain it. Capacity issues aside, there may be other reasons for us to worry that this content could not in principle be contained in anything that resembles the traditional lexicon.

This is the issue I address here, considering two coupled questions. First, What representational content must be ascribed to a word? and second, What representational mechanism is best suited for that content?

The answers I suggest arise from two different strands of research. The first involves computational simulations of phenomena that were not in fact specifically designed to address questions of lexical representation. Rather, their focus was on trying to understand how a neural

network might handle (if indeed, it could at all) sentence level phenomena such as long distance dependences and hierarchical structure. An unanticipated outcome of this research was to suggest a novel way of thinking about words and about the lexicon. The precise implications of this alternative way of representing lexical knowledge were not evident at the time, however.

The second strand of research involved empirical investigations into human sentence processing. Most of this work has been in collaboration with Mary Hare and Ken McRae. Our goal has been to study expectancy generation in sentence processing. What information and what mechanisms are used to help a comprehender anticipate upcoming words of an incrementally presented sentence? Verbs are particularly interesting in this regard because they play a particularly important role in binding together sentential elements, and they impose specific constraints on the arguments and structures with which they occur. Our working assumption was that if we discovered verb-specific factors that influenced expectancy generation, these factors would need to be included in the verb's lexical representation. Many such factors have been discovered (by other researchers as well as ourselves). Some of these might plausibly be placed in the lexicon. However, an important conclusion of this work has been that these factors are bound together by event knowledge. Furthermore, we found that a comprehender's knowledge of events plays a central role in sentence processing, that this knowledge interacts with structural interpretation at the earliest possible moment. This knowledge is not readily incorporated into the lexicon, but as I will argue, there is also not an obvious principled basis for excluding it. This then raises the question: Is it possible to have lexical knowledge without a lexicon? Surprisingly, the early computational work turned out to suggest an answer for the problem posed by the later empirical work.

This is also the order in which I describe these two strands of work: The computational studies (the solution) first, followed by the empirical studies (the problem). I do so because although the solution provided by the computational work will not become apparent until after the experimental data are described, the computational framework has many properties that motivated much of the experimental work. The computational research has been described at length elsewhere in the literature, so will be described only briefly here. The bulk of what follows will focus on the experimental results and why they are problematic for the lexicon.

2. The problem of time

Connectionist models of the early 1980s provided an exciting new computational framework for understanding a number of important phenomena in human behavior for which symbolic serial processing seemed ill adapted. These included the role of context in perception and action, the parallel processing of information, and the ability to rapidly integrate information from multiple sources. But human behavior also unfolds over time, and the architectures of early connectionist models did not deal with temporal processing in a very satisfactory way. Various proposals have been advanced since then to address that shortcoming. I focus here on one class of models that involves the use of recurrent connections (Elman, 1990; Jordan, 1986). These connections give the network access to its own state at prior points in time, thus giving it a kind of memory. The work described below involves one such architecture, known as a Simple Recurrent Network (or SRN; Elman, 1990), shown in **Figure 1**.

-- Insert Figure 1 here –

Recurrent networks can be trained to process time series of various sorts (sequences of phonemes, words, articulatory gestures, etc.) using a simple but powerful learning algorithm (Rumelhart, Hinton, & Williams, 1986). Training is example-based, meaning the network is presented with many examples of inputs and outputs. The goal, however, is to discover a single set of network parameters that allows the network not only to produce the correct output, given (**Figure 1**)the input, but to generalize to novel stimuli. Thus, the training data are used to discover the underlying function that has generated them. What the network is trained to do (i.e., what the target output be should be, given the input) depends on the task. Many tasks have been used with recurrent networks. One simple but very powerful task is prediction.

Prediction is appealing for a number of reasons. For one thing, the information needed for teaching is an observable (i.e., whatever actually occurs next). Thus, no magic oracles are required and everything necessary for learning is available from the environment. Doing prediction can of course be extremely challenging. If the temporal structure of the time series is complex (e.g., involves recursion or long distance dependencies), successful learning requires that the network discover that underlying structure. From an ecological point of view, prediction is highly adaptive. It is easy to imagine many situations in which being able to anticipate the future state of the world, given the present state, is crucial. Finally, there is considerable empirical evidence for prediction in language (e.g., Altmann & Kamide, 2007; DeLong, Urbach, & Kutas, 2005; Kamide, Altmann, & Haywood, 2003; Pickering & Garrod, 2007; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) as well as in other realms of behavior and in the brain (Dayan, 2002; Kochukhova & Gredeback, 2007).

Recurrent networks turn out to have a number of properties that are relevant for language learning. Given an unsegmented sequence of inputs (acoustic sounds, or orthographic

characters), a network will learn to make context-dependent predictions that approximate the

conditional probabilities of succeeding elements. Since conditional probabilities across

boundaries tend to be less constrained, this leads to increases in prediction error at those

boundaries. This provides powerful evidence for implicit boundaries that might not be explicitly

marked in the input stream. It is likely that human infants use a similar mechanism of statistically

driven induction to learn word boundaries in continuous speech (Saffran, Aslin, & Newport,

1996) as well as patterns of a more rule-like nature (Gerken, 2006, 2007; Gomez & Gerken,

2000; Newport & Aslin, 2004). Such learning seems to be possible in nonlinguistic modalities

(Creel, Newport, & Aslin, 2004) and in nonhumans as well (Newport, Hauser, Spaepen, & Aslin,

2004; Pons, 2006).

SRNs can also be trained on sentences, presented a word at a time. In this case, the

distributional restrictions on the contexts in which words occur cause the network to learn

internal representations that reflect both grammatical categories and lexico-semantic information.

This information is encoded in the network's hidden layer by employing a spatial encoding to

position similar elements and categories close in the representational space. When these

representations are analyzed by hierarchical clustering, they reveal a pattern such as the one

shown in **Figure 2**.

-- Insert Figure 2 about here –

A particularly important finding is that SRNs can encode long distance dependences of

the sort that that arise in questions or in sentences with embedded clauses. The fact that the form

of the first occurrence of the verb *to be* in the sentence *The girl whose parents are on vacation in*

Paris is staying with my friend is in the plural (*are*), whereas the second occurrence is in the

singular (*is*) is a puzzle from the viewpoint of surface linear order (because the first occurrence

agrees with the second noun, and the second occurrence agrees with the first noun). In the

sentence *The book I bought cost a lot of money*, the direct object of *buy* occurs not where it might

normally be expected, following the verb, but instead appears at the beginning of the sentence,

leaving a gap and resulting in *buy* being immediately followed by another verb, *cost.* SRNs not

only learn such regularities but generalize to novel (and more complex) cases (e.g., Boden &

Blair, 2003; Christiansen & Chater, 1999; Elman, 1991; Elman, 1993; Lewis & Elman, 2001).

How is this done? The networks cannot simply be memorizing the training data, given their

ability generalize to more complex cases.

　　　To answer this, Rodriguez, Wiles, and Elman (1999) studied networks that had been

trained on one of the simplest formal languages from the class of context-free languages. Such

languages have well known computational requirements that also appear in natural languages.

The language learned by the networks was $a^n b^n$. In this language, sentences consist only of the

words (or terminals) *a* and *b*, and are grammatical only in strings in which some number *n* of

sentence initial *a*s is followed by the same number *n* of *b*s. *aaabbbb, ab, aaaaaabbbbbb,* are

grammatical; *b, a, ba,* and *aaabb* are not. By simplifying the language (while still remaining in

the class of context-free languages), a network of the form show in Figure 1 could be used, but

with only two input and output units, and only two hidden and two context units. This made it

possible to study the network's dynamical structure to see how the dynamics might be harnessed

to perform language relevant computation.

　　　These dynamics depend on the recurrent connections. One way to visualize the dynamics

is to look at the forces that these connections create on the two dimensional internal state of the

network (the two dimensions correspond to the 0.0 to 1.0 activation ranges of the two hidden

units). These dynamics are depicted in Figure 3 and **Figure 4**. These figures show the hidden

unit state space, the state the network is in when a given input is received (i.e., the hidden units'

activations resulting from that input), and—marked by arrows—the vector flow field. This flow

field indicates where successive states will be, assuming the same input. When the flow field

converges toward a fixed point in the space, that marks an attractor. In the case of divergence

away from a fixed point, that marks a repellor. Because the dynamics depend in part on the input,

which acts as a bias on the system, the same network can have multiple regimes. In this case,

because there are two inputs, the network has two different regimes. One occurs when an *a* is

input; the other occurs with *b* inputs.

When an initial *a* is presented to the network, the state moves from its start position

(shown as *S* in Figure 3) to the position pointed to by the trajectory labeled *a1*. In this regime, the

flow field corresponds to an attracting fixed point, and subsequent *a*s result in network states that

oscillate around the attractor, getting closer to it with each *a*. When the first *b* is received, it does

two things. The *b* alters the network's dynamics so that there is now a repelling fixed point.

Second, the *b* moves the network's state to the position shown by the *b1* trajectory. This is

shown in Figure 4. Succeeding *b*s cause the state to oscillate away from the fixed point until it

finally returns to the start state. The network's expectations (predictions about next input) are

determined by its state at any given point in time. When that state is to the left of the partition

shown in Figures 3 and 4, the network expects *a*; when it is to the right of the partition, it

predicts *b*. The key to the solution is that the rate of contraction toward the attracting fixed point

is the inverse of the rate of expansion during the repelling fixed point regime. The number of

initial *a*s determines how close the network gets to the attractor; this also determines how far the

state will be from the repellor when the first *b* comes in, and thus how many more *b*s it will take

for the network to return to the start state.

-- Insert Figures 3 and 4 about here –

This solution turns out to generalize to a number of more complex languages, including

palindrome languages (e.g., characterized by sentences such as ACBDDBCA; Rodriguez, 2001;

Rodriguez & Elman, 1999). Given finite precision, the solution allows the network to generalize

beyond the complexity of the sentences on which it has been trained, with a gradual decline in

performance as length increases. Interestingly, the network shows the same sensitivity in its

performance as do humans to an interaction between depth of embedding and the content of the

material that is embedded. Triply embedded sentences containing semantically similar elements,

such as *The mouse the cat the dog saw chased ran away* are quite difficult, compared to the

relatively easier *Do you believe the report the stuff they put in Coke causes cancer is true?*

(Weckerly & Elman, 1992).

2.2 Language as a dynamical system

The perspective that language processing takes place within a dynamical system, rather

than a symbolic framework, leads to a different way of thinking about rules and words. Rule-like

behavior is achieved through the system's dynamics. A single network may be capable of

supporting multiple dynamical regimes, because in addition to perturbing the network's states, an

input may also function as a bias that changes the dynamics. Collectively, these multiple

dynamical regimes encode a grammar. The grammaticality of a given utterance is then reflected

by the degree to which the sequence of words it is composed of produce trajectories through the system's state space that are consistent with the dynamics.

The role of words in such a dynamical system is to function as external stimuli that alter the system's internal state. The effect that a given word produces is a function of two things: The prior state of the network, which encodes the context in which word input occurs; and the network's dynamical structure or grammar, which is encoded in its weights.

In this scheme of things there is no data structure that corresponds to a lexicon. There are no lexical entries. Rather, there is a grammar on which words operate. Crucially, the system has the capacity to reflect generalizations that occur at multiple levels of granularity. The dynamics may be sensitive to a word's grammatical category, the many conceptual categories it may belong to, and even its specific identity.

Obviously, although the information that one might place in a lexicon is now shifted into the network's dynamics, that same information must still be accounted for even if it is in a different way. Thus, we are offered the possibility of lexical knowledge without a lexicon. What is unclear at this point is what benefit this might bring, if any. Is there any reason to prefer this conceptualization of words over the traditional view?

In what follows, I argue that there is indeed a set of phenomena for which this dynamical account of words offers a more satisfying account than the traditional lexicon.

3. Sentence processing and the lexicon

Within the psycholinguistic literature, much of the data that motivate an enriched lexicon come not from the direct study of lexical representations per se, but from a theoretical debate regarding the mechanisms of sentence processing. The controversy has to do with how language

users deal with the challenge of interpreting sentences that are presented in real time, incrementally, word by word. In many cases, the partially presented fragments may be at least temporarily ambiguous in the sense that they are compatible with very different grammatical structures and very different meaning interpretations. Usually (but not always), the ambiguities are eventually resolved by the remainder of the sentence. The question is how comprehenders deal with the temporary ambiguities at the point where they arise. Two major possibilities have been proposed. Both assume that the comprehender does something, that is, neither assumes that processing is suspended until the end of the sentence. The theories differ on how the comprehender deals with the ambiguity when it occurs.

The historically earlier hypothesis was that processing occurs in at least two stages (e.g., Frazier, 1978, 1990; Frazier, 1995; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). Two-stage theories are motivated by assumptions regarding limitations in human working memory and processing capacity. These limitations force reliance on a number of syntactic heuristics in order to make a provisional parse of a sentence as it is being processed.

During the first stage, the comprehender attempts to create a syntactic parse tree that best matches the input up to that point. It is assumed that, initially, only basic syntactic information regarding the current word is available, such as the word's grammatical category and a limited set of grammatically relevant features. In the case of verbs, this information might include the verb's selectional restrictions, subcategorization information, thematic roles, etc. (Chomsky, 1965, 1981; Dowty, 1991; Katz & Fodor, 1963).

At a slightly later point in time, a second stage of processing occurs in which fuller information about the lexical item becomes available (including the word's semantic and pragmatic information, as well as world knowledge). Interpretive processes also operate, and

these may draw on contextual information. Occasionally, the information that becomes available during the second pass might force a revision of the initial parse. However, if the heuristics are efficient and well motivated, this two-stage approach permits a quick and dirty analysis that will work most of the time without the need for revision.

The contrasting theory, often described as a constraint-based, probabilistic, or expectation-driven approach, emphasizes the probabilistic and context-sensitive aspects of sentence processing (Altmann, 1998, 1999; Altmann & Kamide, 1999; Elman, Hare, & McRae, 2005; Ford, Bresnan, & Kaplan, 1982; MacDonald, 1993; MacDonald, Pearlmutter, & Seidenberg, 1994; MacWhinney & Bates, 1989; McRae, Spivey-Knowlton, & Tanenhaus, 1998; St.John & McClelland, 1990; Tanenhaus & Carlson, 1989; Trueswell, Tanenhaus, & Garnsey, 1994). This approach assumes that comprehenders use all idiosyncratic lexical, semantic, and pragmatic information about each incoming word to determine a provisional analysis. Of course, temporary ambiguities in the input may still arise, and later information in the sentence might reveal that the initial analysis was wrong. Thus, both approaches need to deal with the problem of ambiguity resolution. The question is whether they make different predictions about processing that can be tested experimentally. [1]

---

[1] It might seem that testing the predictions of the two approaches would be straightforward, since one could look to see when a comprehender's behavior is affected by information in the input, relative to when it was available. However, although constraint-based, probabilistic models assume that all information enters the processing stream as soon as it is available in the input, this does not mean that the information will have an instantaneous effect. Constraints differ in their strength, and constraint-based systems often have complex dynamics that arise from competition or synergies between different constraints. Exactly when any given

This has led to a fruitful line of research that focuses on cases in which a sentence is temporarily ambiguous and allows for two (or more) different structural interpretations. What is of interest is what happens when the ambiguity is resolved and it becomes clear which of the earlier possible interpretations is correct. The assumption is that if the sentence is disambiguated to reveal a different structure than the comprehender had assumed, there will be some impact on processing, either through an increased load resulting from recovery and reinterpretation, or perhaps simply as a result of a failed expectation. Various measures have been used as markers of the processing effect that occurs at the disambiguation point in time, including reading times, patterns of eye movements, or EEG activity. These measures in turn provide evidence for how the comprehender interpreted the earlier fragment and therefore (a) what information was available at that time and (b) what processing strategy was used. Clearly, the many links in this chain form a valid argument only when all the links are well motivated; if any aspect of the argument is faulty, then the entire conclusion is undermined. It is not surprising that this issue has been so difficult to resolve to everyone's satisfaction.

Over the years, however, the evidence in favor of the constraint-based, probabilistic approach has grown, leading many (myself included) to view this as the more viable approach. In what follows, I begin by describing several studies in which the results imply a great deal of detailed and verb-specific information is available to comprehenders. The first set of data are

---

piece of information has a measurable effect on the comprehender's overt behavior will depend on the time it takes for these dynamics to be resolved. Nonetheless, there are timing differences between constraint-based and two-stage approaches that can be modeled, cf. (McRae et al., 1998).

amenable to the strategy of an enriched lexicon. I then turn to data for which this is a much less reasonable alternative. These are the data that pose a dilemma for the lexicon.

3.1. Meaning as a cue to structure

One 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*. In this context, *the story* (at the point where it occurs) 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 sentence). The two-stage model predicts that the DO interpretation will be favored initially, even though *hear* admits both possibilities, and there is support for this prediction (Frazier & Rayner, 1982). However, proponents of the constraint-based approach have pointed out that at least three other factors might be responsible for such a result: (1) the relative frequency that a given verb occurs with either a DO or SC (Garnsey, Pearlmutter, Meyers, & Lotocky, 1997; Holmes, 1987; Mitchell & Holmes, 1985); (2) the relative frequency that a given verb takes a SC with or without the disambiguating but optional complementizer *that* (Trueswell, Tanenhaus, & Kello, 1993); and (3) the plausibility of the postverbal NP as a DO for that particular verb (Garnsey et al., 1997; Pickering & Traxler, 1998; Schmauder & 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., Ferreira & Henderson, 1990; Mitchell, 1987). 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 then 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 & 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 (2004; Hare, McRae, & Elman, 2003) 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 (till 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, Elman, Tabaczynski, & McRae, this volume) 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 in Elman, McRae, and Hare (2005).

A similar demonstration of the use of meaning to predict structure is reported in Hare, Elman, Tabaczynski, and McRae (this volume). 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 not simply include 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 results in a slight enrichment of the verb's lexical representation, but can be easily accommodated within the traditional lexicon.

3.2 Verb-specificity of thematic roles

Another well studied ambiguity is that which 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. 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 found innocent.*)

In an earlier study, Taraban and McClelland (1988) found that when participants read sentences involving ambiguous prepositional attachments, e.g., *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,* etc. 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.

There is one additional finding that provides an important qualification of this conclusion. It turns out that different adjectival modifiers of the same noun can also affect its inferred thematic role. Thus, a *shrewd, heartless gambler* is a better agent of *manipulate* than a *young, naïve gambler*; conversely, the latter is a better filler of the same verb's patient role (McRae, Ferretti, & Amyote, 1997).If conceptually-based thematic role preferences are verb-specific, the preferences seem to be finer grained than simply specifying the favored lexical items that fill the role. Rather, the preferences may be expressed at the level of the features and properties that characterize the nominal.

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 et al., 1998) results further suggest a level of information that goes considerably beyond the limited set of proto-role features envisioned by Dowty. (This is reminiscent of McCawley's (1968) observations regarding the use of Katz and Fodor (1963) binary features to encode verb selectional restrictions. McCawley pointed out that verbs such as *devein* and *diagonalize* would require positing selectional features such as [+shrimp] and [+matrix].) 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.

Do verb-specific preferences for their thematic role fillers arise only in the course of sentence processing? Or might such preferences also be revealed in word-word priming? The answer is yes. Ferretti, McRae, and Hatherell (2001) found that verbs primed nouns that were good fillers for their agent, patient, or instrument roles. In a subsequent study, McRae, Hare, Elman, and Ferretti (2005) tested the possibility that such priming might go in the opposite direction, i.e., that when a comprehender encounters a noun, the noun serves as a cue for the event in which it most typically participates, thereby priming verbs that describe that event activity. This prediction is consistent with literature on the multiple forms of organization of autobiographical event memory (Anderson & Conway, 1997; Brown & Schopflocher, 1998) (Lancaster & Barsalou, 1997) (Reiser, Black, & Abelson, 1985). As predicted, priming was found.

The above experiments further 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 studies reviewed so far suggest that the lexical representation for a verb such as *admit* would include subentries about all the verb's senses. For each sense, all possible subcategorization frames would be shown, with information regarding the probability of each, and verb-specific information for each argument/thematic role would appear. The experimental evidence indicates that in many cases, role filler specifications will be detailed and at couched at the featural level (e.g., Ferretti et al., 2001; McRae et al., 1997). Similar constraints would have

to be included frames subcategorization frames. For example, the 'concede, acknowledge' sense of *admit* requires either a patient or proposition that is subject to question or doubt; to *admit* entails the possibility of *deny* (cf. Hare, McRae, & Elman, 2003). This information must also be reflected in the lexical representation. If different information is required for every sense of the verb, and senses are tried to event knowledge, one must be prepared to accommodate a proliferation of such entries.

3.3 Flies in the ointment

Thus far, the experimental data suggest that comprehenders' knowledge of fairly specific (and sometimes idiosyncratic) aspects of a verb's usage is available and utilized early in sentence processing. This information includes sense-specific subcategorization usage patterns, as well as the properties of the nominals that are expected to fill the verb's thematic roles. All of this expands the contents of the verb's lexical representation, but not infeasibly so. Now we come to another set of phenomenon that will be problematic for the traditional view of lexical representation.

3.3.1 Fly #1: The effect of aspect

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? One possibility is that locations are not as tightly associated with an event as are other participating elements. However, 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 mark

resultative information or states that follow the concluded event (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 on-going; 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.

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. Consistent findings were also found in another study in which Event Related Potentials (ERPs) were recorded, indicating that aspect exerted its effect early in processing.

A similar effect of aspect on sentence processing has recently been found in the quite different domain of pronoun interpretation. This research is concerned with how people interpret a personal pronoun in one sentence when there are two potential referents in a previous sentence, and both are of the same gender (e.g., *Sue disliked Lisa intensely. She_____*). In this case, the reference is ambiguous.

One possibility is that there is a fixed preference, such that the pronoun is usually construed as referring to the referent that is in (for example) Subject position of the previous sentence. Another possibility, suggested by Kehler, Kertz, Rohde, and Elman (2008) is that pronoun interpretation depends on the inferred coherence relations between the two sentences (Kehler, 2002). Under different discourse conditions, different interpretations might be preferred.

In a prior experiment, Stevenson, Crawley, and Kleinman (1994) asked participants to complete sentence pairs such as *John handed a book to Bob. He___* in which the pronoun could equally refer to either *John* (who in this context is said to fill the Source thematic role) or Bob

(who fills the Goal role). Stevenson et al. found that Goal continuations (in which *he* is understood as referring to *Bob*) and Source continuations (*he* refers to *John*) were about evenly split, 49%-51%. Kehler et al. suggested that, as was found in the Ferretti et al. (2007) study, aspect might matter. The reasoning was that perfective aspect tends to focus on the end state of an event, whereas imperfective aspect makes the on-going event more salient. When the event is construed as completed, the coherence of the discourse is most naturally maintained by continuing the story, what Kehler (2002) and Hobbs (1990) have called an Occasion coherence relation. Because continuations naturally focus on the Goal, the preference for Goal interpretations should increase. This appears to be the case. When participants were given sentences in which the verb was in the imperfective, such as *John was handing a book to Bob,* and then asked to complete a following sentence that began *He___*, participants generated significantly more Source interpretations (70%) than for sentences in which the verb had perfective aspect. This result is consistent with the Ferretti et al. (2007) interpretation of their data, namely, that aspect alters the way comprehenders construe the event structure underlying an utterance. This in turn makes certain event participants more or less salient.

Let us return now to the effect of aspect on verb argument expectations. These results have two important implications. First, the modulating effect of aspect is not easily accommodated by spreading activation accounts of priming. In spreading activation models, priming is accomplished via links that connect related words and which serve to pass activation from one to another. These links are not thought to be subject to dynamic reconfiguration or context-sensitive modulation. In Section 4, I describe an alternative mechanism that might account for these effects.

The second implication has to do with how verb argument preferences are encoded. 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 Event Related Potential methodology indicated aspectual differences within 400 ms of the expected word). The immediate accessibility and impact of this information would make it a likely candidate for inclusion in the verb's lexical representation. But logically, it is difficult to see how one would encode such a dynamic contingency on thematic role requirements.

Thus, although the patterns of ambiguity resolution described in earlier sections, along with parallel findings using priming (Ferretti et al., 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 a similar account. A verb's aspect is not an intrinsic property of the verb, yet the particular choice of aspect used in a given context affects expectations regarding the expectations regarding the verb's arguments.

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 several of the earlier studies. 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.

3.3.2 Fly #2: Different agents, different instruments: Different events?

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 that 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 question is, What is the status of such knowledge? No one doubts that a comprehender's

knowledge of how and what a surgeon cuts, versus a lumberjack, plays an important role in

comprehension at some point. The more critical issue is when this knowledge is brought to bear,

because timing has implications for models of processing and representation. If the knowledge is

available very early—perhaps even immediately on encountering the relevant cues—then this is

a challenge for two-stage serial theories (in which only limited lexical information is available

during the first stage). Importantly, it is also problematic for standard theories of the lexicon.

3.3.2.1 *Agent effects*

Bicknell, Elman, Hare, McRae, and Kutas (in preparation) 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 and 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. This was done because we were interested in event-based effects of the agent that are mediated by the verb. 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, there was an increase in reading times for sentences in which an agent-verb combination was followed by an incongruent (though plausible) patient. The slowdown occurred 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. More to the point, the effect occurred at the

first possible moment in processing.

3.3.2.2. Instrument effects

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 correspond 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, Elman, Hare, and McRae (in preparation) 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 a self-paced reading format, participants read 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 normed to

ensure that there were no direct lexical associations between instrument and patient, and that

instruments in isolation did not differentially prime the patients. 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.

3.3.3. Fly #4: Situational effects

The problems for traditional lexical representations should start to be apparent. But there is one final twist. So far, we have seen that expectations regarding one of a verb's arguments may be affected by how another of its arguments is realized. Is this effect limited to argument-argument interactions, or can discourse level context modulate argument expectations?

Race, Klein, Hare, and Tanenhaus (in preparation) took a subset of the sentences used in the Bicknell et al. experiment, in which different agent-verb combinations led to different predictions of the most likely patient. Race et al. then created stories that preceded the sentences, and in which the overall context strongly suggested a specific event that would involve actions that might or might not be typical for a given agent-verb combination. For example, although normally *The shopper saved...* and *The lifeguard saved...* lead to expectations of some amount of *money* and some of *person*, respectively, if the prior context indicates that there is a disaster occurring, or if there is a sale in progress, then this information might override the typical expectancies. That is exactly what Race et al. found. This leads to a final observation: A verb's preferred patients do not depend solely on the verb, nor on the specific filler of the agent role, nor on the filler of the instrument role, but also on information from the broader discourse.

Now let us see what all of this implies as far as the lexicon is concerned.

4. Lexical knowledge without a lexicon

These data suggest three lessons regarding the factors that influence expectancy generation during sentence processing and priming. First, comprehenders are sensitive to ways in

which the syntactic structures that are expected for a given verb depend on the context-specific sense in which the verb is used. Second, comprehenders are sensitive to contingencies between the specific fillers of a verb's agent and instrument roles, on the one hand, and the expected filler of the verb's patient role. Third, these contingencies are also subject to broader contextual effects which may include discourse or verbal aspect, so that altering the context can give rise to different contingencies.

In a superficial sense, none of this is surprising, since clearly these factors are known to comprehenders and ought to affect comprehension at some stage in processing. What make the data significant are two things.

First, the timing of the effects—specifically, the fact that they occur immediately at the first possible moment in time—is not consistent with two-stage serial accounts in which a syntactic analysis precedes, and is uninformed by, semantic, pragmatic, or world knowledge. This is essentially the position outlined in 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" (p. 435). More pithily, the data do not accord with the "syntax proposes, semantics disposes" hypothesis (Crain & Steedman, 1985). There are now many findings of this sort in the literature, and these data add to them.

The second question, and the one I take as more interesting, concerns what the data suggest about lexical representation. The earlier sets of data that have been reviewed are part of a much larger body of findings from experiments devised to test predictions of the constraint-based, probabilistic framework. Cumulatively, all of these data are usually interpreted as

indicating a far more central role for the lexicon in sentence processing than was initially

envisioned, and they also suggest that lexical representations contain a significant amount of

detailed word-specific information that is available and used during on-line sentence processing.

However, problems arise when we move from claiming that there is verb-specific

information regarding preferred thematic role fillers to claiming that (a) this information depends

on things such as the particular aspect with which the verb is used; (b) the information regarding

preferred fillers of one argument or thematic role depends on what the filler is of one of the other

roles (in a way suggested, for example, by the contingencies on patient expectations for the verb

cut, depicted in **Figure 5**; and (c) everything depends on other qualifying information, stated or

implied, in the discourse regarding the nature of the event being described by the verb. The

difficulty lies not simply in the combinatoric explosion entailed by having to encode such

contingencies; it is difficult to envision how the potentially unbounded number of contexts that

might be relevant could be anticipated and stored in the lexicon.

-- Insert Figure 5 about here –

So why would one even think of putting this sort of information into the lexicon? This

raises the more fundamental question: What criteria should we use to decide what goes into the

lexicon and what does not?

This question has no easy answer. There is considerable controversy regarding what sort

of information belongs in the lexicon, with different theories taking different and often mutually

incompatible positions (contrast, among many other examples, Chomsky, 1965; Haiman, 1980;

Jackendoff, 1983, 2002; Katz & Fodor, 1963; Lakoff, 1971; Langacker, 1987; Levin & Hovav,

2005; Weinreich, 1962). The challenge presented by the sort of data presented above (and indeed, there are many other findings that point in the same direction) is that if all of the factors that these data indicate are relevant do indeed participate in generating verb argument expectancies, this leaves us with three general possibilities for deciding where this information belongs.

1. Only information that is systematic and generalizes across classes of verbs should be included in a verb's lexical representation; or

2. All of the information that is relevant to verb argument possibilities should be included in its lexical entry; or

3. None of this information belongs in the lexical entry.

The first position is the tidiest and has the appeal of avoiding the Scylla and Charybdis posed by the other two options. It preserves the time-honored (though not uncontroversial) distinction between lexical semantic information, which belongs in the lexicon, and world knowledge. Assuming a parallel architecture in which different language modules have the opportunity for full and bidirectional interaction (e.g., Jackendoff, 2002), this position would accommodate the present data.

The devil is in the details. Whether or not this is a feasible alternative depends on how one defines systematic, and what processes are assumed to underlie generalization. This question seemed less thorny earlier in time when the range of data that were considered to be informative was restricted to grammaticality judgments (and even then, only from trained linguists). But increasingly, linguistic analyses rely what were once dismissed as 'merely' performance data; these include reading times, analyses of usage in large text corpora, patterns of eye movements, child and adult language acquisition data, data from aphasic patients, and neuroimaging. Notions

such as systematicity, productivity, and idiosyncrasy have turned out to be largely a matter of degree. In my view, a brutal but honest assessment is that this solution—as appealing as it might be for its tidiness—is probably not tenable.

The second option represents the logical conclusion of the trend that has appeared not only in the processing literature (e.g., in addition to the studies cited above, Altmann & Kamide, 2007; Kamide, Altmann et al., 2003; Kamide, Scheepers, & Altmann, 2003; van Berkum et al., 2005; van Berkum, Zwitserlood, Hagoort, & Brown, 2003) but also many recent linguistic theories (e.g., Bresnan, 2006; Fauconnier & 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?

The main problem here is this. If we allow the lexicon to contain any and all information that is relevant to the use and interpretation of a word, and this information can come from any and all knowledge sources, is it plausible that the identical information exists both in the lexicon and in these other sources? What purpose would be served? One good answer (which I agree with) is that linguistic forms, while relating to conceptual and world knowledge, are also subject to constraints that are specific to the linguistic domain. This is perfectly true. Reading the word *onion* and smelling an onion are not the same thing. Seeing a horse fall after running past a barn is not the same thing as hearing *The horse raced past the barn fell.*

The question is whether these facts require a separate copy of a person's conceptual and world knowledge, plus the additional facts that are specific to linguistic forms. Or is there some other way by which such constraints can operate directly on a shared representation of

conceptual and world knowledge? In more prosaic terms, can we take the world out of language and put language in the world?

4.1 Modeling event schemas

Language can be used for many purposes. Let us assume, however, that a great deal of language involves reference to the knowledge that language users have about events and situations. These events constitute a common ground between interlocutors, and any given discourse may build on this knowledge to serve the particular needs of the interactions. In any given sentence, the linguistic elements serve as cues that help the comprehender access that knowledge, and also encourage specific construals.

The verb is one powerful constraint on which event type is being referred to, but the participants themselves may also constrain the event. Some cues are more potent than others. S*ee, person,* or *yesterday* are very weak cues as to the underlying event. On the other hand, knowing that the activity is *diagonalizing* allows only one possible patient (and tells us something about the mathematical knowledge of the agent). We hear a sentence that begins *She kneeled and put her head in the guillotine…* and we know not only what will happen next, but whose head is about to roll, as well as something about the time and location of the event. Some cues are more necessary than others. A verb is almost always required, though *You, me, and a bottle of wine* requires little imagination to infer the hoped for event. Finally, multiple cues may interact, such that their value together is different than their values apart. Langacker's notion of 'accommodation' and Pustejovsky's examples of 'enriched composition' are instances of this (Langacker, 1987; Pustejovsky, 1996; see also McElree, Traxler, Pickering, Seely, & Jackendoff,

2001). *Finish the book* means very different things in the sentences *The author finished the book, The student finished the book, My goat finished the book.*

The similarity between event knowledge and schema theory should be apparent. The schema (and its relatives, the frame, story, and script Abelson, 1981; Minsky, 1997; Norman & Rumelhart, 1981; Schank & Abelson, 1988) figured prominently in the psychological literature of the 1970s because it seemed to provide a powerful mechanism for organizing our knowledge about many common activities. Unfortunately, the computational architectures that were available at the time for implementing schema had an undesirable degree of brittleness and inflexibility. But this was a property of the architectures that were used, not of the concept itself. Today there exist computational architectures that in fact seem to have characteristics that are better suited for instantiating the schema. Connectionist models, for example, have properties that make it possible to implement constraints that are graded, and which allow context-sensitive filling of roles. A simple but elegant demonstration of this appears in a model of room schemas that was developed by Rumelhart, Smolensky, McClelland, and Hinton (1988).

The Rumelhart et al. model consists of an interconnected network of 40 nodes, in which each node stands for a room descriptor (e.g., *walls, dresser, carpet,* scale). Subjects were asked to imagine different types of rooms and to describe them, drawing from the 40 room descriptors. The strength of connections between any two nodes was then adjusted to reflect the cooccurrence statistics of those room descriptors. As a result, if the nodes for *oven* and *ceiling* are turned on, this leads to gradual coactivation of other nodes that collectively described a kitchen. Turning on the nodes for *bed* and *ceiling* leads to activation of nodes for room elements that would be found in a bedroom. Furthermore, it is possible to turn on nodes for items in combinations that do not appear in any typical room, but which lead to a sensible blend of room types. When this happens,

some typical properties of rooms change to accommodate the blend. For instance, turning on *bed* and *sofa* leads to a state that might be described as a large, fancy bedroom.

In this model, elements function as constraints. Some are weak, others are strong. There are default values for some features, but these may be altered in response to context. Schemas themselves are epiphenomenal. They emerge as a result of the patterns of (possibly higher order) cooccurrence among the various participants in the schema. The same network may instantiate multiple schemas, and different schemas may blend. Finally, schemas emerge as generalizations across multiple individual examples. Although this particular model did not involve learning, this could be implemented.

These are just the sort of properties one wants of an event model, with the important addition of a temporal dimension. Events unfold over time, and causes precede their results. Similarly, a model of the language used to describe an event requires a temporal dimension, since sentences are interpreted word by word. What might such a model look like? How might it behave? Here I return to the recurrent network architecture shown in Figure 1.

A very simplified version of such a model was trained on a corpus of sentences that described a variety of events involving different agents, instruments, and patients. After training, sentences were presented, one word at a time, to examine the network states that result and the trajectory of these states over time. **Figure 6** shows the response to the two probe sentences *A person uses a saw to cut a tree* and *A butcher uses a saw to cut meat*. Only three of the 20 dimensions (in the network's hidden unit space) are shown. The state of the network that is produced by a word, in context, is what encodes the expectancies for what will follow. Thus *cut* with a *saw* with a generic *person* agent generates different expectancies than does *cut* with a *saw* with *butcher* as agent.

-- Insert Figure 6 about here –

Variants of the same event type (e.g., *cutting*) follow similar paths, but the differences reflect the ways in which the specifics of the event differ depending on participants (in this case, different instruments leading to expectations of different patients). Different event types (e.g., *eating* vs. *cutting*) are represented by different families of trajectories. The families of trajectories might be thought of as instantiating a general event schema, and event variants as corresponding to specific subtypes of events within that family.

The model in Figure 7 is far too simple to serve as anything but a conceptual metaphor. It is intended to help visualize how the knowledge that we are removing from word-as-operand is moved into the processing mechanism on which word-as-operator acts. The model also illustrates one way to understand the suggestion by Rumelhart (quoted in the epigraph to this paper) that words should be thought of not as having intrinsic meaning, but as providing cues to meaning.

Critically, this simple model is disembodied; it lacks the conceptual knowledge about events that comes from direct experience. The model only captures the dynamics of the linguistic input. In a full model, one would want many inputs, corresponding to the multiple modalities in which we experience the world. The dynamics of such a system would be considerably more complex than those shown in Figure 7, since each input domain has its own properties and domain internal dynamics. In a more complete model, these would exist as coupled dynamical subsystems.

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 (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 (they are conventionalized). (Fauconnier, 1997; Fauconnier & Turner, 2002; Hutchins, 1994)

(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). (Kintsch, 1988; Langacker, 1987; van Berkum et al., 2005; van Berkum et al., 2003)

(3) The drive to predict is a simple behavior with enormously important consequences. It is a powerful engine for learning, and provides important clues to latent abstract structure (as in language). Prediction lays the groundwork for learning about causation. (Elman, 1990; Kahneman & Tversky, 1973; Schultz, Dayan, & Montague, 1997; Spirtes, Glymour, & Scheines, 2000)

(4) Events play a major role in organizing our experience. Event knowledge is used to drive inference, to access memory, and 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. (Minsky, 1997; Schank & Abelson, 1988; Shipley & Zacks, 2008; Zacks & Tversky, 2001)

(5) 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. (Smith & Thelen, 2003; Spencer & Schöner, 2003; Spivey & Dale, 2004; Tabor, Juliano, & Tanenhaus, 1997; Thelen & Smith, 1994)

Nonetheless, doing away with lexicon is radical surgery. What are the implications of the alternative that is suggested here?

5.1 Domain specificity?

Does viewing language in this broader context of cognition and action mean that language has no independent life? Is the claim that language makes no specific contribution to cognition? Aren't there regularities and facts that are specific and unique to language alone?

No, no, and yes.

All stimuli within a modality possess properties that to varying degrees are specific and perhaps even unique to that modality. The processing of those stimuli requires sensitivity to those properties (typically the subset that is important to the perceiver). Insofar as those stimuli reflect facts about their generator, we can consider them as adhering to a grammar. In some domains the grammars are primarily determined by physical properties of the world. In social and cultural domains, the grammars tend to be primarily conventional and so depend on shared habits of usage. The point here is merely that although the words-as-cues perspective assumes tighter coupling between linguistic and nonlinguistic processes, it also assumes that the linguistic stream will possess characteristics that are unique to its domain.

Furthermore, it is also clear that although the word *onion* may tap directly into nonlinguistic knowledge of onions, the state that is evoked is not the same as when one sees an onion, or when one is smelled. These stimuli access the information in different ways. Nor is it solely a matter of access. Language is constructive as well as evocative. For example, we can describe situations that do not exist (*Imagine now a purple cow*), or which could not exist (*Colorless green ideas sleep furiously*). In such cases, we are drawing on experience but language allows us to use those experiences in imaginatively new ways. When our imagination falters, the sentence does not become meaningless (contra Chomsky, 1965); rather, the meaning is simply bizarre.

5.2 How does business change?

Although I have argued that much of the behavioral phenomena described above are not easily incorporated into the lexicon, I cannot at this point claim that accommodating them in some variant of the lexicon is impossible. 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, could probably account for the data described earlier. Concerns would remain about how to motivate what information is placed where, but these concerns do not in themselves rule out a lexical solution. Unfortunately, it is also then not obvious whether tests can be devised to distinguish between these proposals.

However, theories can also be evaluated for their ability to offer new ways of thinking about old problems, or to provoke new questions that would not be otherwise asked. A theory might be preferred over another because it leads to a research program that is more productive

than the alternative. Let me suggest two positive consequences to the sort of words-as-cues dynamical model I am outlining.

The first has to do with the role that theories play in the phenomena they predict. 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, the notion that selectional restrictions might be dynamic and context-sensitive is fundamentally not an option within the Katz and Fodor framework (1963). 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 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 been longer in the coming than it might have, given a different theoretical perspective.

It is also curious that even within the field of language research, there is compartmentalization in the study of phenomena. I believe this too is a consequence of the highly modular theories of language and cognition that prevailed for decades. Syntactic ambiguity resolution, pronoun interpretation, priming, text inference, and semantic memory (to chose but a small subset of domains) are typically studied as unrelated phenomena, and the respective literatures rarely make contact with one another. Yet these phenomena—as the studies reviewed here illustrate—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 et al., 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. And of course, the discourse context may provide further constraints on how an event is construed. Thus, as Race et al. 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.

There is a second consequence to viewing linguistic and nonlinguistic cues as tightly coupled. This has to do with learning and the problem of learnability. Much has been made about the so-called poverty of the stimulus (Chomsky, 1980, p. 34; Crain, 1991). The claim is that the linguistic data that are available to the child are insufficient to account for certain things that the child eventually knows about language. Two interesting things can be said about this claim. First, the argument typically is advanced "in principle" with scant empirical evidence that it truly is a problem. A search of the literature reveals a surprisingly small number of specific phenomena for which the poverty of the stimulus is alleged. Second, whether or not the stimuli

available for learning are impoverished depend crucially on what one considers to be the relevant and available stimuli, and what the relevant and available aspects or properties of those stimuli are.

Our beliefs about what children hear seem to be based partly on intuition, partly on very small corpora, and partly on limited attempts to see whether children are in fact prone to make errors in the face of limited data. In at least some cases, more careful examination of the data and of what children do and can learn given those data do not support the poverty of the stimulus claim (Ambridge, Pine, Rowland, & Young, 2008; Pullum & Scholz, 2002; Reali & Christiansen, 2005; Scholz & Pullum, 2002). It is not always necessary to see X in the input to know that X is true. It may be that Y and Z logically make X necessary (Lewis & Elman, 2001).

If anything is impoverished, it is not the stimuli but our appreciation for how rich the stimuli the stimuli are. The usual assumption is that the relevant stimuli consist of the words a child hears, and some of the arguments that have been used in support of the poverty of the stimulus hypothesis (e.g., Gold, 1967) have to do with what are essentially problems in learning syntactic patterns from positive only data. We have no idea how easy or difficult language learning is if the data include not only the linguistic input but the simultaneous stream of nonlinguistic information that accompanies it. However, there are many examples that demonstrate that learning in one modality can be facilitated by use of information from another modality (e.g., Ballard & Brown, 1993; de Sa, 2004; de Sa & Ballard, 1998). Why should this not also be true for language learning as well?

*   *   *

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 is brought to bear on the way language is processed.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*(7), 715-729.

Altmann, G. T. M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences, 2*, 146-152.

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory & Language, 41*(1), 124-145.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure. *Cognition, 106*(1), 87-129.

Anderson, S. J., & Conway, M. A. (1997). Representations of autobiographical memories. In M. A. Conway (Ed.), *Cognitive Models of Memory* (pp. 217-246). Cambridge, MA: MIT Press.

Ballard, D. H., & Brown, C. M. (1993). Principles of animate vision. In *Aloimonos, Yiannis (1993). Active perception. Computer vision. (pp. 245-282). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc. viii, 292 pp.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes, 12*, 507-584.

Boden, M., & Blair, A. (2003). Learning the dynamics of embedded clauses. *Applied Intelligence, 19*(1-2), 51-63.

Bresnan, J. (2006). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter.

Brown, N. R., & Schopflocher, D. (1998). Event clusters: An organization of personal events in autobiographical memory. *Psychological Science, 9*, 470-475.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Oxford, England: M.I.T. Press.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1981). *Lectures on Government and Binding*. New York: Foris.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*(2), 157-205.

Crain, S. (1991). Language acquisition in the absence of experience. *Brain and Behavioral Sciences, 14*, 597-611.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parserarsw. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives* (pp. 320-358). Cambridge: Cambridge University Press.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1119-1130.

Dayan, P. (2002). Matters temporal. *Trends in Cognitive Sciences, 6*(3), 105-106.

de Sa, V. R. (2004). Sensory modality segregation. In S. Thurn, L. Saul & B. Schoelkopf (Eds.), *Advances in Neural Information Processing Systems* (Vol. 16, pp. 913-920). Cambridge, MA: MIT Press.

de Sa, V. R., & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation, 10*(5), 1097-1117.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*, 547-619.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science, 14*(2), 179-211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195-224.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99.

Elman, J. L., Hare, M., & McRae, K. (2005). Cues, Constraints, and Competition in Sentence Processing. In *Tomasello, Michael; Slobin, Dan Isaac (2005). Beyond nature-nurture: Essays in honor of Elizabeth Bates. (pp. 111-138). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. il, 339 pp.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Fauconnier, G. (1997). *Mappings in thought and language*. New York, NY: Cambridge University Press.

Fauconnier, G., & Turner, M. (2002). *The way we think : conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*(4), 555-568.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb Aspect and the Activation of Event Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 182-196.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*(44), 516-547.

Fodor, J. D. (1995). Thematic roles and modularity. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 434-456). Cambridge, MA: MIT Press.

Ford, M., Bresnan, J., & Kaplan, R. M. (1982). A Competence-Based Theory of Syntactic Closure. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations* (pp. 727-796). Cambridge: MIT Press.

Frazier, L. (1978). *On comprehending sentences: syntactic parsing strategies.* Unpublished PhD, University of Connecticut.

Frazier, L. (1990). *Parsing modifiers: Special-purpose routines in the human sentence processing mechanism*. Hillsdale, NJ: Erlbaum.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research Special Issue: Sentence processing: I, 24*(6), 437-468.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.

Garnsey, S. M., Pearlmutter, N. J., Meyers, E., & Lotocky, M. A. (1997). The contribution of verb-bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58-93.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*(3), B67-B74.

Gerken, L. (2007). Acquiring linguistic structure. In *Hoff, Erika; Shatz, Marilyn (2007). Blackwell handbook of language development. Blackwell handbooks of developmental psychology. (pp. 173-190). Malden, MA, US: Blackwell Publishing. xii, 502 pp.* Malden, MA: Blackwell Publishing.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 16*, 447-474.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences, 7*(5), 219-224.

Gomez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science, 4*(5), 178-186.

Haiman, J. (1980). Dictionaires and encyclopedias. *Lingua, 50*, 329-357.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization prefences. *Journal of Memory & Language, 48*(2), 281-303.

Hobbs, J. R. (1990). *Literature and cognition*. Stanford, CA: Center for the Study of Language and Information.

Holmes, V. M. (1987). Syntactic parsing: In search of the garden path. In M. Coltheart & et al. (Eds.), *Attention and performance 12: The psychology of reading* (pp. 587-599). Hove, England UK: Lawrence Erlbaum Associates, Inc, 1987 xvi, 707.

Hutchins, E. (1994). *Cognition in the World*. Cambridge, MA: MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, and Evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Res, 1146*, 2-22.

Jordan, M. I. (1986). Serial order: A parallel distributed processing approach. *Institute for Cognitive Science Report 8604. University of California, San Diego.*, -.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37-55.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language, 39*(2), 170-210.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, Calif.: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics, 25*, 1-44.

Kennison, S. M. (1999). American English usage frequencies for noun phrase and tensed sentence complement-taking verbs. *Journal of Psycholinguistic Research, 28*(2), 165-177.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Kochukhova, O., & Gredeback, G. (2007). Learning about occlusion: Initial assumptions and rapid adjustments. *Cognition, 105*(1), 26-46.

Lakoff, G. (1971). Presuppositions and relative grammaticality. In D. Steinberg & L. Jakobovitz (Eds.), *Semantics*. London: Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things : what categories reveal about the mind*. Chicago: University of Chicago Press.

Lancaster, J. S., & Barsalou, L. (1997). Multiple organisations of events in memory. *Memory5, , 5*, 569-599.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar* (Vol. 1). Stanford: Stanford University Press.

Levin, B., & Hovav, M. R. (2005). *Argument realization*. Cambridge ; New York: Cambridge University Press.

Lewis, J. D., & Elman, J. L. (2001). A connectionist investigation of linguistic arguments from the poverty of the stimulus: Learning the unlearnable. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 552-557). Mahwah, NJ: Lawrence Erlbaum Associates.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory & Language, 32*(5), 692-715.

MacDonald, M. C. (1997). Lexical representations and sentence processing: An introduction. *Language & Cognitive Processes, 12*(2&3), 121-136.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703.

MacWhinney, B., & Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing*. Cambridge; New York: Cambridge University Press.

McCawley, J. D. (1968). The role of semantics in a grammar. In E. Bach & R. T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 124-169.). New York: Holt, RInehart and WInston.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition, 78*(1), B17-B25.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing, 12*, 137-176.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition, 33*(7), 1174-1184.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory & Language, 38*(3), 283-312.

Minsky, M. (1997). A framework for representing knowledge.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 601-618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local ambiguity. *Journal of Memory and Language, 24*, 542-559.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*(2), 85-117.

Norman, D. A., & Rumelhart, D. E. (1981). The LNR approach to human information processing. *Cognition, 10*(1), 235-240.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105-110.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden-paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 940-961.

Pons, F. (2006). The Effects of Distributional Learning on Rats' Sensitivity to Phonetic Information. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(1), 97-101.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review, 19*(1-2), 9-50.

Pustejovsky, J. (1996). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 358-374.

Reali, F., & Christiansen, M. H. (2005). Uncovering the Richness of the Stimulus: Structure Dependence and Indirect Statistical Evidence. *Cognitive Science: A Multidisciplinary Journal, 29*(6), 1007-1028.

Reiser, B. J., Black, J. B., & Abelson, R. P. (1985). Knowledge structures in the organization and retrieval of autobiographical memory. *Cognitive Psychology, 17*, 89-137.

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation, 13*(9), 2093-2118.

Rodriguez, P., & Elman, J. L. (1999). Watching the transients: viewing a simple recurrent network as a limited counter. *Behaviormetrika, 26, 1*, 51-74.

Rodriguez, P., Wiles, J., & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science, 11*(1), 5-40.

Roland, D., & Jurafsky, D. (1998). *How verb subcategorization frequencies are affected by corpus choice.* Paper presented at the Proceedings of COLING-ACL 1998, Montreal, Canada.

Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo & S. Stevenson (Eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*: John Benjamins.

Rumelhart, D. E. (1979). Some problems with the notion that words have literal meanings. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 71-82). Cambridge: Cambridge University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning Internal Representations by Error Propagation*. 2, Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1988). Schemata and sequential thought processes in PDP models.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning By 8-Month-Old Infants. *Science, 274*(5294), 1926-1928.

Schank, R. C., & Abelson, R. P. (1988). Scripts, plans, goals, and understanding.

Schmauder, A. R., & Egan, M. C. (1998). The influence of semantic fit on on-line sentence processing. *Memory & Cognition, 26*(6), 1304-1312.

Scholz, B. C., & Pullum, G. K. (2002). Searching for arguments to support linguistic nativism. *Linguistic Review, 19*(1-2), 185-223.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Shipley, T. F., & Zacks, J. M. (Eds.). (2008). *Understanding Events: How Humans See, Represent, and Act on Events*. Oxford: Oxford University Press.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 343-348.

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6*(4), 392-412.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.

Spivey, M. J., & Dale, R. (Eds.). (2004). *On the continuity of mind: Toward a dynamical account of cognition*. San Diego, CA: Elsevier Academic Press.

St.John, M., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence, 46*, 217-257.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9*(4), 519-548.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language & Cognitive Processes, 12*(2&3), 211-271.

Tanenhaus, M. K., & Carlson, G. (1989). *Lexical structure and language comprehension*. Cambridge, MA: MIT Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Taraban, R., & McClelland, J. L. (1988). Constituent Attachment and Thematic Role Assignment in Sentence Processing - Influences of Content-Based Expectations. *Journal of Memory and Language, 27*(6), 597-632.

Thelen, E., & Smith, L. B. (1994). *A dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Science, 4*, 156-163.

Tomasello, M. (2003). *Constructing a language: A Usage-Based Theory of Language Acqusition*. Cambridge, MA: Harvard University Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory & Language, 33*(3), 285-318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*(3), 528-553.

van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(3), 443-467.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*(3), 701-718.

Weckerly, J., & Elman, J. L. (1992). *A PDP approach to processing center-embedded sentences*. Hillsdale, NJ: Erlbaum.

Weinreich, U. (1962). Lexicographic definition in descriptive semantics. In F. W. Householder & W. Saporta (Eds.), *Problems in lexicography* (Vol. 21). Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Figure Captions

Figure 1. Simple Recurrent Network. Each layer is composed of one or more units. Information flows from input to hidden to output layers. In addition, at every time step t, the hidden unit layer receives input from the context layer, which stores the hidden unit activations from time t-1.

Figure 2. Hierarchical clustering diagram of hidden unit activation patterns in response to different words. The similarity between words and groups of words is reflected in the tree structure; items that are closer are joined lower in the tree.

Figure 3. Network dynamics and changes in hidden unit state space while processing a inputs. The axes correspond to activation states for Hidden Units 1 and 2. The heavy solid line indicates the a/b decision plane; network states to the left of the plane result in output predictions that the next element will be an a; states to the right of the plane result in b predictions. Arrows indicate the vector flow field (network dynamics). The network begins and ends in a state within the region marked S. In this figure and the next figure, the trajectory of network states produced by the sequence aaabbb is shown. As each input is received, it shifts the state from the previous position to a new position; a1 shows the first of these transitions. Successive as lead to oscillations around a fixed point attractor whose dynamics are indicated by the vector flow field (small arrows in the background). When the first b is received, it moves the network to a region that is to the right of the a/b decision plane, and alters the dynamical regime so that successive bs oscillate away from a fixed point repellor. The rate of contraction around the attractor is the
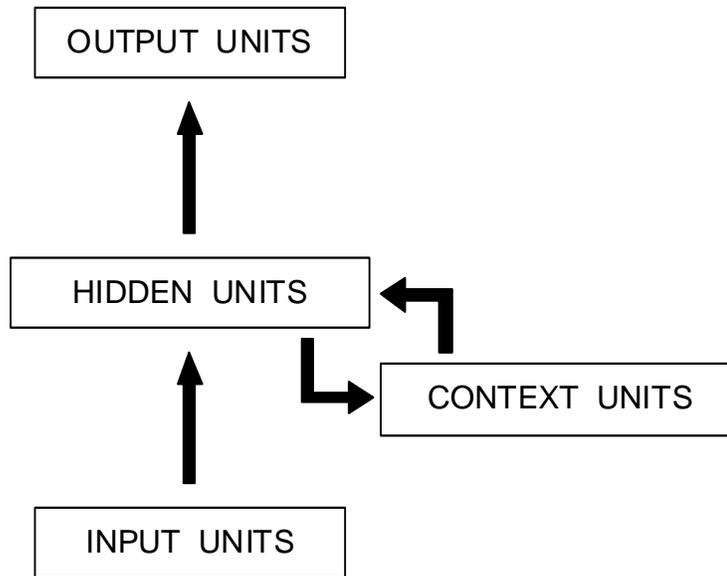
inverse of the rate of expansion around the repellor. The location of the final a determines the

location of the initial b, thus guaranteeing that just as many bs will be needed as there were as in

order to return to the starting region, which is on the a side of the a/b decision plane.
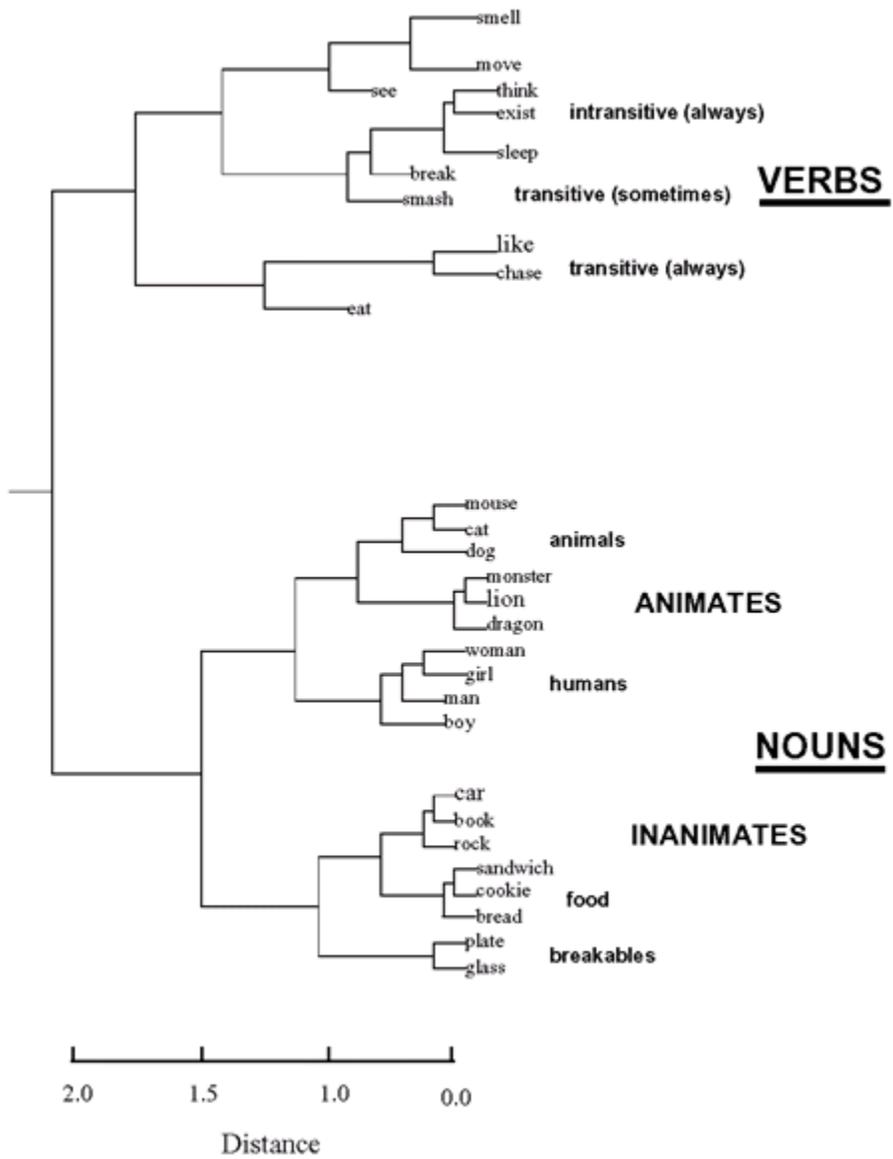
Figure 4. Network dynamics and changes in hidden unit state space while processing b

inputs. In this figure and the next figure, the trajectory of network states produced by the b

portion of the sequence aaabbb is shown. The location of the last input (a3, shown in the

previous figure) determines the location of the initial b. Because of the inverse relationship

between the A regime attractor and the B regime repellor, this ensures that 3 bs will be required

for the network to return to the start state, which then leads to predicting that the next input will
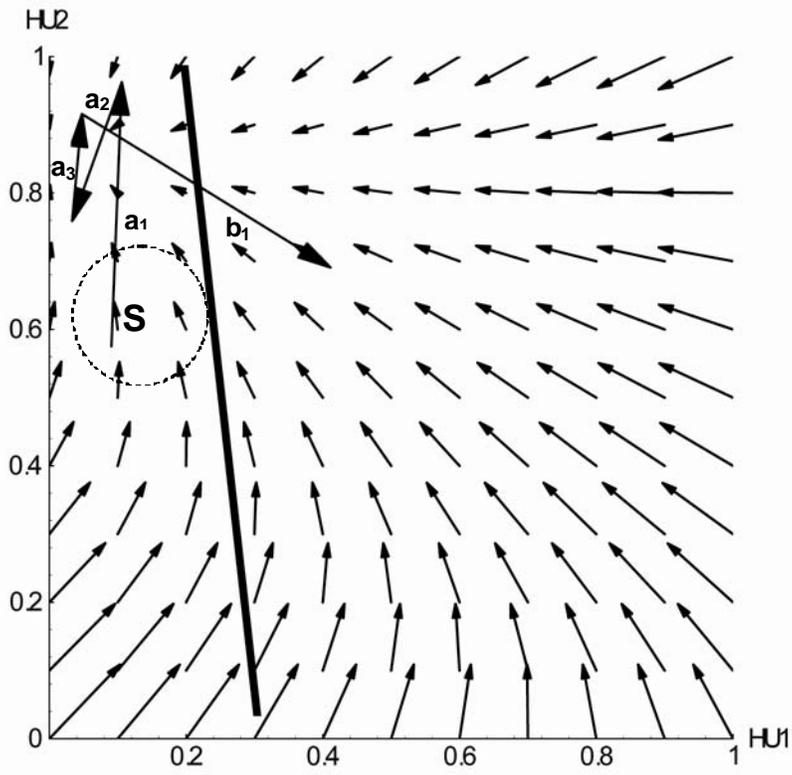
be the a signaling a new sentence.

Figure 5. A schematic illustration of some of the contingent factors that affect

expectations for the patient of cut. The expectations may differ, depending on the identify of the

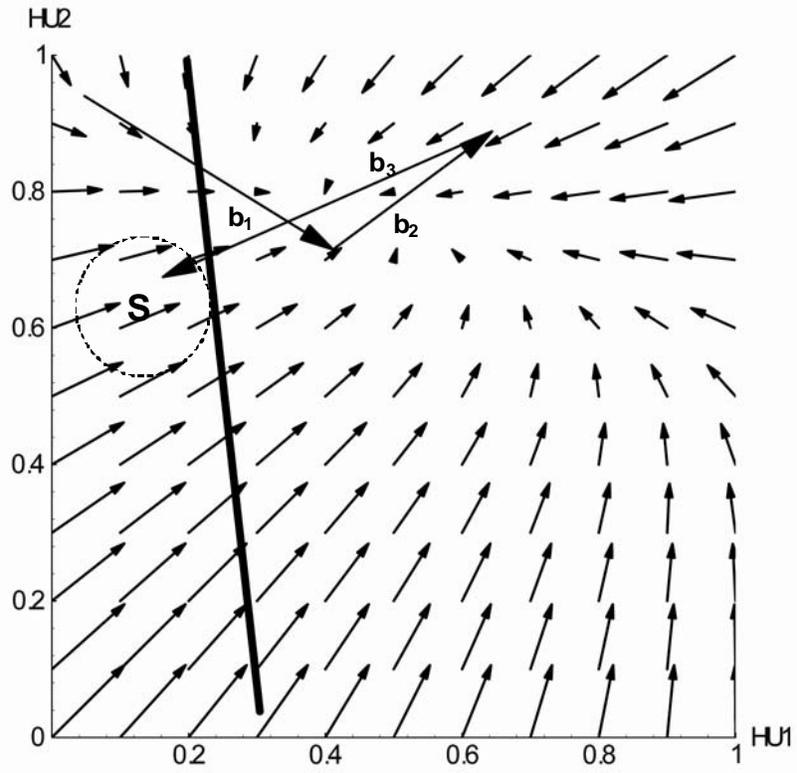agent, the instrument, or the location of the underlying event.

Figure 6. Trajectories through 3 (of 20 total) dimensions of an SRN's hidden layer. These

correspond to movement through the state space as the network processes the sentences "A

person uses a saw to cut a tree" and "A butcher uses a saw to cut meat." The state of the network

resulting from any given word is what encodes its expectancies of what will follow. Thus, the

states at "cut" in the two sentences differ, reflecting different expectations regarding the likely

patient this is to follow (resulting from the use of different instruments). Once the patient is
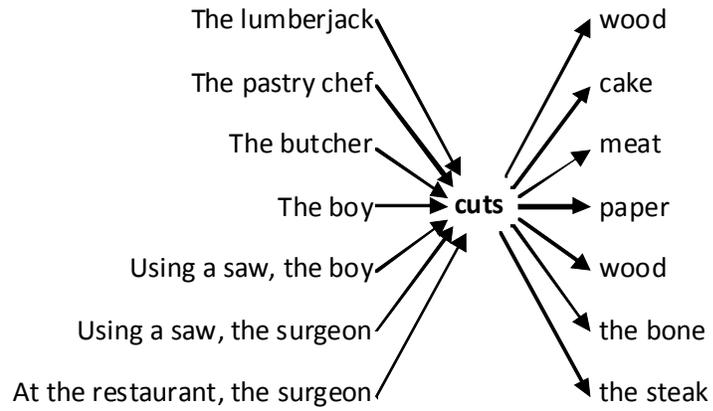
processed, it produces a state appropriate to the end of the sentence. This is why both patients
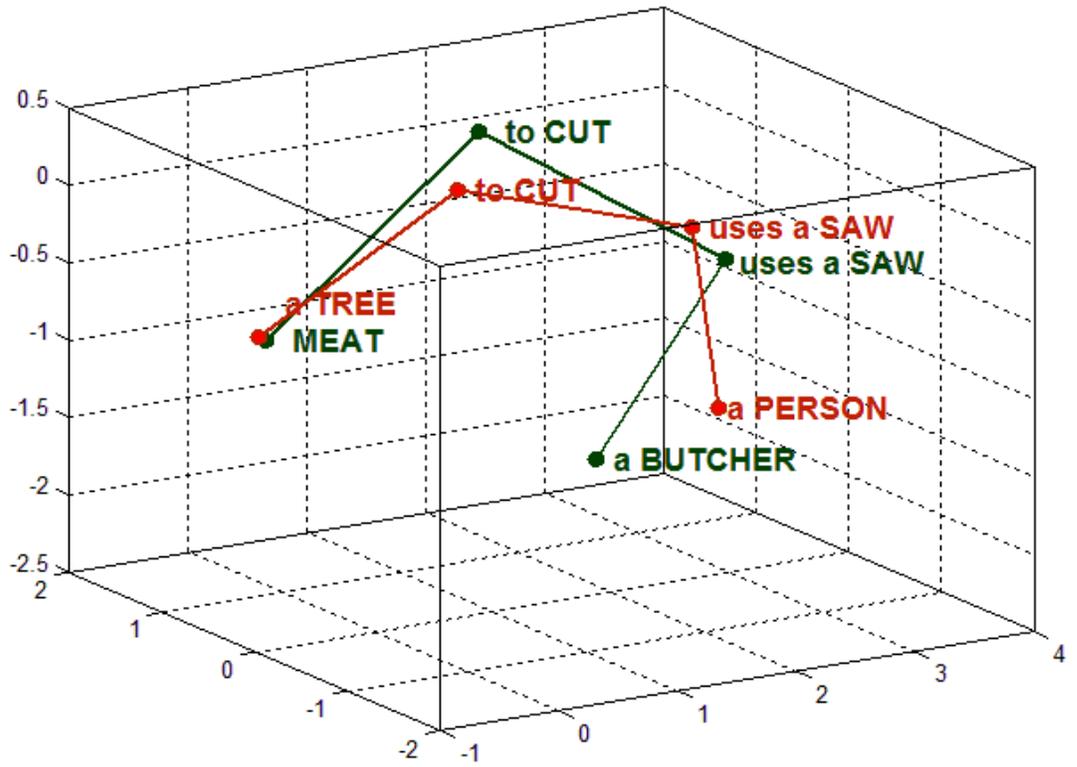
produce very similar states.

```
                    ┌─────────────────┐
                    │  OUTPUT  UNITS  │
                    └─────────────────┘
                             ▲
                             │
                             │
         ┌─────────────────┐         ┌──────────────────┐
         │  HIDDEN  UNITS  │◀────────│  CONTEXT  UNITS  │
         └─────────────────┘────────▶└──────────────────┘
                    ▲
                    │
                    │
         ┌─────────────────┐
         │  INPUT  UNITS   │
         └─────────────────┘
```

The lumberjack → cuts → wood
The pastry chef → cuts → cake
The butcher → cuts → meat
The boy → cuts → paper
Using a saw, the boy → cuts → wood
Using a saw, the surgeon → cuts → the bone
At the restaurant, the surgeon → cuts → the steak

Author Notes