On The Meaning of Words and Dinosaur Bones:
Lexical Knowledge Without a Lexicon

Jeffrey L. Elman
Department of Communication
University of California, San Diego
La Jolla, CA

Correspondence concerning this article should be addressed to:

Jeff Elman
Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0526
Email: jelfman@ad.ucsd.edu

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. Interestingly, the studies reviewed have some methodological and theoretical affinities with General Semantics, a theory of human nature, knowledge, and language from the early 20th century. The turn to the study of words and knowledge in language processing, I suggest, could fruitfully return to this period of inquiry for inspiration. The results of the studies reviewed pose 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] lists a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics.
Ray Jackendoff

Introduction

In this paper, I propose sweep away a common concept of linguistic theories and its assumptions about the nature of human knowledge and meaning. The arguments I develop will not only try to eliminate this concept (i.e., *the mental lexicon*), but also unsettle the assumptions about language and knowledge that have led to this concept's great success. I want to suggest the possibility of lexical knowledge and language processing *without* a lexicon. It is not my goal to call into question the existence of words, or the many things that language users know about them and, by connection, the world. Instead, I wish to question the presuppositions by which lexical knowledge is encoded as the content of words and also question current common wisdom about the representational mechanisms that are best suited to operate on this lexical knowledge.

As a contribution to this symposium, I wish to draw some suggestive methodological affinities between language research and General Semantics. Although it has for some time been out of favor in scientific circles, General Semantics is relevant to recent debates in linguistic research and its assumptions about the nature of knowledge, the mediation of behavior, and the way we process language. Methodologically, the relationship between linguistic comprehension and "semantic evaluation" (a key concept in General Semantics), and the relationship between words and mutli-ordinality, could be very interesting to explore further. Thus in addition to reviewing the lines of research, theoretical advances, and experimental results that led to the reflections in this paper, I will try to mark where substantive connections may be drawn between

contemporary linguistics and General Semantics in the hopes that a backward glance may help chart a way forward.

Rules, Words, and Lexical-Knowledge

For many years, a sharp distinction was made in language research between rules and words -- with a primary focus on rules. The apparently idiosyncratic character of words (in the sense that the relationship between word-meaning and word-form is arbitrary and varies randomly across languages) made words relatively uninteresting for many language researchers. The focus was rather on deciphering the rules of grammar. This distinction, however, is now blurred in many theories of language. In fact, emphasis has recently shifted attention in favor of words. Many linguists have come to see words not simply as the flesh that gives life to grammatical structures, but as bones that are themselves grammatically 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).

In short, a growing number of results from many different areas of language research suggest that lexical representations -- i.e., words -- 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 language processing. These results have led to the emerging consensus that a great deal of information and knowledge is represented by words. The kind and quality of this information varies across these studies. However, the general conclusion has been that words are mental representations that contain information. This information can change, but the information required to parse (comprehend; evaluate) word or sentence meaning, it is argued, is sufficiently represented by -- which in this case means *contained in* -- the words themselves.

Representation is not only a question of content, but also of form, and this form is central to arguments about active language processing. What representational mechanism is required to operationalize the argument that lexical-knowledge is encoded as the content of words? Here, there are many positions, but the proposal offered by Jackendoff in the epigram above seems to reflect a rough consensus among many language researchers: lexical-knowledge is stored as the discrete entries (words) of a mental dictionary (a *lexicon*). The precise form of the lexicon varies according to different theories, but almost all theories assume that the lexicon is an enumerative data structure with some principled constraints on the nature of the information that may be stored within it. In short, the turn to the study of words to understand language comprehension and production has fostered, or perhaps required, the belief that words are mental containers encoding lexical knowledge, and that, either consequently or apriori, words are organized by the mechanism of a mental lexicon.

As the amount of lexical-knowledge presumably contained by words increases, one might start to wonder whether lexical content exceeds either the capacity of a lexical item or, indeed, the form of the lexicon itself. I take this to be interesting, but in fact a minor concern. A more serious problem would arise if lexical knowledge were shown to be context-dependent rather than self-sufficient, self-contained, and relatively static. Small dependencies between lexical-knowledge and context-dependent knowledge might be tolerated, but if the combinatorics of context effects increase (as they are wont to do), a strictly enumerative data structure (e.g., a dictionary) seems infeasible. I argue here that this is in fact the case. I will try to dispute the assumption that the best way to understand language processing is by assuming that words are discrete, lexical items that contain lexical knowledge, and that a mental lexicon organizes all of these lexical items in an enumerative data structure. Faced with the inherent plurality of human experience and the radical heterogeneity of contexts in which experience is conditioned, an enumerative data structure cannot adequately account either for the form of lexical-knowledge or for the language processing mechanism required to account for the many phenomena of human language. I will present some research that concludes that neither the brain nor language processes function in this manner.

The Nature of Language and General Semantics

The emphasis on words as static encodings of lexical knowledge indexed in a mental lexicon depends, in part, on a broader set of assumptions about the nature of language that have been dominant since the 1950s. At the core of these assumptions lies the belief that the capacity for language is an innate, biological function of an area of the brain, entirely unconnected with the contexts of experience. The early linguists' attention to the rules of grammar sought to discover

the universal rules by which syntax could, so to speak, take care of semantics. Chomsky, a major figure of this period, argued that language comprehension and production unfolds from rules to which a speaker's bodily, culturally and historically contingent experience is irrelevant; so too the speaker's particular individual cognitive and psychological development. In short, the early study of language, but also human behavior, required a metaphysical account of the nature of language completely divested from the structures of individual experience, development, and communication with others. It is true, of course, that the shift in the study of language from rules to words has also entailed a shift away from theories emphasizing grammar over vocabulary. But what has largely remained in place is the assumption that the mind's capacity for language is still best understood via the structured, static anatomy of the brain (to wit, Jackendoff's mental lexicon is a "long-term memory store," harking to the cognitive psychology of the 1950s). Such enduring assumptions have made it difficult for the recent turn in linguistics to words to cash out on the notion that lexical-knowledge might be context-dependent, embodied, historically and culturally mediated.

It is with an eye to proposing a different view that we can usefully draw on General Semantics. *Science and Sanity*, Korzybski's most famous work, still bears interesting and relevant lessons for the study of language processes (i.e., *evaluation*, in General Semantics terms), the nature of knowledge in human beings, the ineluctable contingencies of context, and the integration of all these in dynamical, distributed systems. One of his most important lessons is that meaning and knowledge cannot be understood or investigated as stable, fixed conditions of language. General semantics argues that the knowledge required to comprehend or use language depends on forms of knowledge and experience that exist outside of language. *Semantic evaluation* -- a concept developed for the general study of human knowledge and

behavior -- can be understood in this context as *language comprehension*. Korzybski sought to develop a theory of *evaluation* that is contingent on the cultural and historical situatededness of human beings, rather then universal or metaphysical rules or assumptions about human nature. Importantly, he argued that language is not simply an innate capacity but, instead, a cultural-historical artifact that endures through time because it can be embodied and exchanged in speech, print, and other media. I think that the experimental results reported in this paper, its attendant questions and discussions, point to the fecundity of General Semantic's method of investigating "knowing" and "speaking." The concept of a *semantic evaluation*, sometimes also called *semantic reactions*, integrates context and extra-linguistic knowledge in an extremely important way for linguists. Interestingly, as far as the experiments that were conducted for this paper goes, Korzybksi based this theory on major shifts in science that unfolded in the early 20th century, which argued that probabilistic, dynamic, constraint-based models are critical to the operation of our nervous system and the operations of our mind. General Semantics and the inquiry into linguistic representation and processes supported here share this common intellectual history, at the base of which lie some important methodological assumptions.

Some of the methodological principles that Korzybski proposed in his inquiry into knowledge, behavior, and semantic evaluation can be usefully brought into relation with those that informed my experiments, their results, and the conclusions I have drawn from them. Among them is the principle of *time-binding*. Another reflects the insight that words, lexically speaking, are not all of the same form, but rather cultural artifacts; words mediate differently how we know, mean, and think in different contexts. While words *are* indeed critical units of language, the form and content of the lexical-knowledge they are presumed to account for is not enough to understand the nature of language in human communication. We must, to paraphrase a

famous phrase, examine how words function as useful *maps* for individuals navigating a *territory*. This will unveil the larger forms by which words, as cultural artifacts, are context-dependent or, in the language of General Semantics, *multi-ordinal*. Moreover, if we place in question the presumed lexical form, content and representation of words, we will also have to place in question the language processes assumed to operate on words. Thus must look at how words operate relationally within dynamic, distributed mental representations. We must try to shift our assumption about the forms and encodings of *mental representations of words and grammar,* and the form language processes might take.

Two Lines of Research

In an earlier paper (Elman, *Cognitive Science*), I presented two lines of research that converged to reject the two the common arguments about words, knowledge, and meaning described above. The first is that mental representations of words are best understood to *contain* lexical-knowledge that is stored within a mental lexicon. The second is that grammar remains a static corpus of rules that, by indexing the lexical-knowledge of words, plucks out word content and operates on it. The first line of research involved designing computational simulations of linguistic phenomena that were not in fact specifically designed to address questions of lexical knowledge. The focus of this research was rather to understand how a neural network might account for sentence level phenomena such as long distance dependencies and hierarchical structure. One of the important presuppositions of this research is that mental processes are better understood as distributed, connectionist models of small, computational units (mimicking the human nervous system) than as linear, symbolic processing machines. Korzybski, who was living through the revolution of quantum mechanics and witnessing the birth of cybernetics

while he was writing *Science and Sanity,* develops his theory of human epistemology and evaluation through a model of thought that reproduces the operations of the central nervous system. An unanticipated outcome of my research was to suggest a novel way of thinking about words, their function in meaning-making and language processing, and about the lexicon. My contention was that, if words were like the nodes of a neural network, then the network and its processes as a whole could be understood as a grammatical encoding. The neural network approach, although I did not recognize this fact at the time, provided a way of thinking about the mental representations of lexical-knowledge but without the assumption of such a thing as a mental lexicon. Rather than reaffirm the idea that the evaluation of sentences relies on a mental lexicon, I concluded that evaluations depend much more on the mutual, co-dependent relations of words within dynamical systems than on the presumed lexical content of words.

The second strand of research involved empirical investigations into human sentence processing. Our goal was to study expectancy generation in sentence processing. What information and what mechanisms are used to help a language comprehender anticipate, or *predict*, upcoming words when the words are incrementally presented in a sentence? Verbs are particularly interesting in this regard because they play an important role in binding together sentential elements, and they impose specific constraints on the arguments and structures with which they occur. My working assumption was that if I discovered verb-specific factors that influence the generation of word-expectations, then these factors should necessarily be included in the verb's lexical entry in the mental lexicon. Many such factors have been discovered (by other researchers as well as us). Some of these factors 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* -- i.e., knowledge that would not be bound to a lexicon -- rather than the lexicon

itself. Furthermore, we found that a comprehender's knowledge of events plays a central role in sentence processing, and that this knowledge interacts with structural interpretation at the earliest possible moment -- too early, in some cases, to be attributed to the passive content of lexical-entries. In short, it seemed to me that *words* were doing a lot more than simply recalling information from a lexicon. Instead, the specificity of particular words *cued* the expectation of subsequent words and sentence meaning. These words, we found, were culturally and contextually and situationally dependent, so that knowledge about how events unfold, who does what to whom with what, and other *event knowledge*, played a critical role in shifting expectancy generation in sentential processing.

This knowledge is not readily incorporated into the lexicon, but as I argued, there is no obvious principled basis for excluding it. Still, our conclusions are anathema to the view that, although words are critical to grammatical interpretation, both words and grammar are generic structures that function independently of cultural-historical context. The knowledge that could not be incorporated into the lexicon includes context-specific senses in which a verb, for example, is used; or the contingencies between the particularities of a verb's agent or instrument roles and the verb's patient role (as in the verb *cut*); or, finally, this knowledge is subject to broader contextual effects which may include discourse or verbal aspect, so that altering the context can give rise to different contingencies. As Korzybski might have concluded from these results, the knowledge required to evaluate a sentence is not co-extensive with the lexical-knowledge words are presumed to represent. In fact, it is better to understand words not as containers of information, but as *maps* that orient the comprehender in a specific context. As Korzybski argued, the way we "know" in order to "make sense" or "make meaning" relies on forms of knowledge outside of "verbal levels."

These experimental results pose a conundrum for traditional views of the mental lexicon. Either the mental lexicon must be expanded to include factors that do not plausibly appear to belong it; or, virtually all information about word meaning needs to be moved outside of the lexical-content of words, leaving the lexicon entirely impoverished. I suggest a third alternative, which provides a way to account for lexical knowledge without a mental lexicon. This third way requires that we not only change our view of what *words* "are" (perhaps it is better to think about what they do and how they do it) and also how language processing works. The traditional view is that language is processed through a two-stage serial process in which syntactic analysis precedes, and is not informed by, semantic, pragmatic, or world knowledge. This is essentially the position outlined by J. D. Fodor (1995): "We may assume that there is a syntactic 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." But the results of these two lines of research led me to adopt the view that language processing must unfold through time, in context, where it was much easier to see that words are operators in a dynamic meaning-making space where, at best, they orient one to meaning. By "operator" I mean that words shift the states of a neural network (an encoding of grammar) rather than supply a grammar with information from a lexicon.

With this background, we can now move on to draw out affinities between my formulation of the relations of words and lexical knowledge to General Semantics. Alfred Korzybski introduced a series of methodological principles that, although directed at the whole study of man and civilization, took seriously the problem of language and speech, and therefore can usefully be directed to linguistics research. Although General Semantics and my research into language are, in many respects, enormously different and even incomparable, I believe we

share certain methodological assumptions that I think would be useful to draw out -- both for future linguistics research and the continuing interest in General Semantics.

Simple Recurrent Networks as Computational Time-Binders

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 phenomena 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. What I want to suggest is that, although not quite perfectly aligned, connectionist models of neural network processes through time computationally implement what Korzybski called the principle of *time-binding* -- i.e., the transmission and accretion of knowledge through time, which in the case of neural networks is encoded in *connections* of the network itself. It seems to me that a recurrent neural network could be said to *encode semantic evaluations*.

Recurrent networks can be trained to process time series of various sorts and levels (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 set of network parameters that allows the network not only to produce the correct

output, given the input, but to generalize its computation to novel stimuli. The training data are used to discover the underlying function that has generated them. In an important way, the processes in a neural network that train the network are akin to what Korzybski called abstraction processes, which is to say the generation of internal coded representations about the world that, through experience, can shift through time. What a network is trained to do depends on the task. 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. That is, once an expectation is generated, it can be confirmed or disconfirmed by simply seeing whatever actually occurs next in time (much like our everyday experience). At the same time, everything that the network requires to learn is observable.

Recurrent networks turn out to have a number of properties that are relevant for language learning, and 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 (Kahneman & Tversky, 1973; Kveraga, Ghuman, & Bar, 2007; Spirtes, Glymour, & Scheines, 2000) and in the brain (Dayan, 2002; Kochukhova & Gredeback, 2007). Given an unsegmented sequence of inputs (acoustic sounds, or orthographic characters), a network will learn to make context-dependent predictions (we can also think of these as evaluations) that approximate the conditional probabilities of the succeeding elements of a sequence. Recurrent Networks 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.

The perspective that language processing takes place within a dynamical system (i.e., neural networks and semantic evaluations), rather than a symbolic framework, leads to a different way of thinking about rules and words. Rule-like behavior is achieved through the system dynamics of the neural network. A single network may be capable of supporting multiple dynamical regimes, because in addition to perturbing the network's states, an input may also change its dynamics. Collectively, these multiple dynamical regimes encode a grammar. But I have also proposed that we think of a *recurrent* neural network as a model of semantic reactions. The grammaticality (i.e., evaluation) 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 space that are consistent with the dynamics. Tabor (1997; 2001) has made similar points, and in a particularly elegant study (2004) has demonstrated that a dynamical approach accounts for the effects of 'local coherence' on processing, in which partial parses that are syntactically compatible with only a part of the input are constructed, even if these are incompatible with a globally syntactic parse.

What does all of this have to do with the lexicon? The critical insight here is that the role of words in such a dynamical system, like *maps* or *abstractions* in General Semantics, 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. First, of the prior state of the network, which encodes the context in which word input occurs -- this context is akin to what Korzybski called a *semantic environment*, the whole of semantic resources and their relations in which a human being can know the world.

Second, the network's dynamical structure or grammar, which is encoded in its weights, and which we could trace as the *semantic reaction* of the individual in the world.

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? The fact that this approach might offer a novel alternative to the lexicon *qua* data structure is interesting but the more important question is what might be gained. 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.

Sentence Processing and the Lexicon

In the psycholinguistic literature, the data that motivate a belief in an enriched lexicon do not come from the direct study of lexical representations, but as a by-product of a theoretical debate regarding the mechanisms of sentence processing. The debate concerns how language users deal with the challenge of evaluating sentences 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 different grammatical structures and different evaluations of meaning. The question is how comprehenders deal with the temporary ambiguities in the moments they arise. Two major possibilities have been proposed, and both assume the comprehender actively works to resolve ambiguity. The theories differ, however, 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 assume limitations to working memory and processing capacity. These limitations force a 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 in this first stage, 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, and thematic roles. (Chomsky, 1965, 1981; Dowty, 1991; Katz & Fodor, 1963). At a 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 this 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 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 is the approach that builds neural networks and computational approximations of time-binding. It assumes that comprehenders use *all* idiosyncratic lexical, semantic, and pragmatic information about each incoming word in order to determine a provisional analysis. Temporary ambiguities in the input may still arise, and later information in the sentence might reveal that the initial analysis was wrong. Both approaches deal with the problem of ambiguity resolution. The question is whether they make different predictions about processing that can be tested experimentally.

This debate 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) structural interpretations. The question is what happens when the ambiguity is resolved and it becomes clear which of the earlier possible interpretations is correct. If the sentence is disambiguated to reveal a different structure than the comprehender had assumed, then there will be some impact on processing, either through an increased cognitive 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 provide evidence of how earlier fragment is interpreted and therefore (a) what information was available at that time and (b) what processing strategy was used. Over the years, the evidence in favor of the constraint-based, probabilistic

approach has grown, leading many to view it as the better model of human sentence processing. It is this research that has supported the enriched lexicon hypothesis. In what follows, I describe several studies in which the results imply a great deal of detailed and verb-specific information is available to comprehenders. Although first set of data are amenable to the strategy of an enriched lexicon, we quickly come upon data for which this is a much less reasonable alternative. These are the data that pose a dilemma for the lexicon.

The results of experiments described below concern the study of verbs in sentence processing; a complete review of these experiments can be found in another paper (Elman, *Cognitive Science*). In this paper, I would like only to briefly recapitulate the nature and results of these experiments in order to suggest an affinity between what language processing and semantic evaluation. The nature of these experiments is to investigate the *variable effects of verbs* on phrase interpretation. What the heterogeneity of these effects reveal is more than the fact that specific verbs may hold a heterogeneity of lexical knowledge; in fact, they reveal that associating lexical knowledge with a verb itself may be misguided, and that lexical knowledge is much more a function of the way in which different *verbs cue other* forms of knowledge not reducible to a lexical entry. In short, verbs operate as maps that, within dynamical semantic environments, call upon other semantic resources in order to interpret (or evaluate) a phrase *through time*. These results would indicate that the concept of the mental lexicon would be unnecessary in sentence processing, and that the latter would not operate by processes that require a mental lexicon.

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*. *The story* could either be the direct object (DO) of *heard*, or it could be the subject noun of a sentential complement (SC).

Various reported experiments have shown mixed results about which kind of structure is more likely to follow in these ambiguities (Ferreira & Henderson, 1990; Mitchell, 1987; Garnsey et al., 1997; Trueswell et al., 1994; Kennison, 1999). One possible explanation for the discrepant 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, 2002). If verb meaning influences subcategorization, then it is likely context influences verb meaning; and 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. Several other studies have in fact demonstrated that the sense-specific use of the word predicts structure (Hare, McRae, & Elman, 2003, 2004; Hare, Elman, Tabaczynski & McRae, *in press*). The experimental results suggest that the lexical representations 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 easily be accommodated within the traditional lexicon.

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 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,* that reading times were faster in sentences involving more *typical* fillers of the instrument role (in these examples, *broom* rather than *solvent*). This led McRae et al. (1998) to hypothesize -- and conclude experimentally -- 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. 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 experiemental data suggest that the roles contain very detailed information about hte preferred filler of these roles, and that the preferences are verb-specific.

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 -- verbs are able to prime their preferred agents, instruments, and patients. 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.

The Effect of Aspect

Ferretti (2001) found that verbs primed their agents, patients, and instruments, but did not find that they primed the locations in which actions were taking place. One possibility is that location is less strongly associated with events. However, Ferretti, Kutas, and McRae (2007) noted that in that experiment the verb primes for locations were in the past tense, and possibily were 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. 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. Thus Ferretti et al. hypotehsized that although a past perfect verb did not prime its associated location, the same verb in the progressive might do so because of the locations' greater saalience to the unfolding event. This was borne out. Several other studies (Kehler, Kertz, Rohde, and Eleman, 2008; Ferretti, 2007; Stevenson, Crawley, Kleinman, 1994; Kehler, 2002; Hobbs, 1990) have found that verbal aspect manipulates sentence processing by changing the focus to an event description. One particular line of reasoning (Kehler, 2002) was that the perfective aspect tneds to focus on the end state of an event, whereas imperfective aspect makes the on-going event more salient.

These results have two important implications. First, the modulating effect of aspect is not easily accommodated by spreading activation accounts of verb 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.

A 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. 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. 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.

Different agents, different instruments: Different events?

If we think in terms of verbs as Korzybskian *maps* and events as the knowledge they target, then it should be clear that although the verb is obviously a very powerful map, and its aspect may alter the way the event is construed, there are other maps that change the nature of the event or activity associated with the verb.

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). This is also problematic for standard theories of the lexicon.

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.

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.

It is significant is that the variation in patient-expectation as a function of particular agent-verb combinations occurs at the earliest possible moment, at the patient that immediately follows the verb. The timing of such effects has in the past often been taken as indicative of an effect's source. A common assumption has been that immediate effects reflect lexical or 'first-pass' processing, and later effects reflect the use of semantic or pragmatic information. In this study, the agent-verb combinations draw upon comprehenders' world knowledge. The immediacy of the effect would seem to require either that this information must be embedded in the lexicon, or else that world knowledge must be able to interact with lexical knowledge more quickly than has often typically been assumed.

In a similar vein, Elman, Hare, and McRae (in preparation) tested the possibility that the *instrument* used with a verb in a sentence would cue different event schemas, leading to differing expectations regarding the mostly 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. An additional priming study was carried out in which instruments and patients served as prime-target pairs; no significant priming was found between typical instruments and patients (e.g., *scissors-paper*) versus atypical instruments and patients (e.g., *saw-paper;* but priming did occur for a set of additional items that were included as a comparison set). As predicted, readers showed increased reading times for the atypical patient relative to the typical patient. In this study, the effect occurred right at the patient, demonstrating that the filler of the instrument role for a specific verb alters the restrictions on the filler of the patient role. Now let us see what all of this implies as far as the lexicon is concerned.

Encoding 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, on the other. 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, however, this is not surprising, since clearly these factors are known to comprehenders and ought to affect comprehension at some stage in processing. On its face, there seems to be no empirical or normative reason why these factors cannot be assimilated into a mental-lexicon concept.

The problem with the mental lexicon arises 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 lies in the difficulty of envisioning how the potentially unbounded number of contexts that might be relevant could be anticipated and stored in the lexicon. If we allow the lexicon to contain all the information that is relevant to the use and interpretation of a word, and if this information can come from any and all knowledge sources, then 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. They are subject to the behavioral constraints of experience and context in which the comprehender is situated. 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, *and* 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, is it possible

to take the world out of language and put language in the world? The answer is yes, but only if we grapple with whether and how the plurality -- or multi-ordinality -- of language can become a general, contingent, and culturally-historically situated condition of human behavior. This question invites, in short, the view that language belongs to dynamic, recurrent, temporal system of interpretation or, in the language of General Semantics, the processes of *semantic evaluation*.

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 possess regarding events and situations. In other words, let as assume that words are *maps, abstractions* of a similar territory generated from pervious experience. Now the event or moments in which language appears as one of the central phenomena is a semantic environment. This semantic environment constitutes 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, we have seen, is one powerful constraint on which event type is being referred to, but the participants themselves may also constrain the event. Some cues, like some maps, are more potent that others in different semantic environments, and organize different semantic reactions. Verbs, as maps, are so informative that they are usually necessary. However, we have also seen that agents, patients, instruments, and locations -- in short, a whole series of ways of knowing and organizing experience through time -- activate the evaluations with which they typically occur (McRae et al., 2005). The role of categories of words other than verbs in cuing event knowledge is particularly important in constructions and languages in which the verb appears late in the sentence. Finally, multiple kinds of *maps* may interact, such that their value

together is different than their values apart -- indeed, it is the semantic environment that the conditions under which words and their relations orient evaluation and meaning. Langacker's notion of "accommodation" and Pustejovsky's examples of "enriched composition" are instances of such interactions (Langacker, 1987; Pustejovsky, 1996; see also McElree, Traxler, Pickering, Seely, & Jackendoff, 2001). In this model, elements (words) function as constraints, linguistic evaluations and conditional on these constraints, and the processes of a network account for a whole, situated semantic reaction. The *maps* that mediated these semantic reactions may be altered in response to context and experience; the schemas themselves are epiphenomenal. They emerge as a result of the patterns of (possibly higher order) co-occurrence among the various participants in the schema. The same network may instantiate multiple schemas, and different schemas may blend, depending on the semantic environment. 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 how language can be 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? A very simplified version of such a network was trained on a corpus of sentences that described a variety of events involving different agents, instruments, and patients. Critically, this simple model is disembodied; it lacks the conceptual knowledge about events that comes from direct experience. The work described here has emphasized verbal language, and this 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.

Discourse involves many other types of interactions. For example, the work of Clark, Goldin-Meadow, McNeil, and many others makes it clear that language is well and rapidly integrated with gesture (H.H. Clark, 1996; 2003; Goldin-Meadow, 2003; McNeil, 1992, 2005). The dynamics of such a system would be considerably more complex than those we have discussed in this paper, since each input domain has its own properties and domain internal dynamics. In a more complete model, these would exist as coupled dynamical subsystems that interact.

Conclusions

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 (or *maps-as-abstractions* in the terms of General Semantics) 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 smells an onion. 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 at extreme variance with the world as we have experienced it.

Although I have argued that many 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, might be able to 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. This remains an open question for the moment.

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 a search 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 might occur early in processing. Many researchers in the field have already come to this point of view. It is a conclusion that, despite considerable empirical evidence, has been longer in the coming than it might have benn, given a different theoretical perspective.

A second consequence of this perspective is that it encourages a more unified view of phenomena that are often treated (de facto, if not in principle) as unrelated. Syntactic ambiguity resolution, lexical ambiguity resolution, pronoun interpretation, text inference, and semantic memory (to chose but a small subset of domains) are studied by communities that do not always communicate well, and researchers in these areas are not always aware of findings from other areas. Yet these domains have considerable potential for informing each other. That is because, although they ultimately draw on a common conceptual knowledge base, that knowledge base can be accessed in different ways, and this in turn affects what is accessed. Consider how our

knowledge of events might be tapped using a priming paradigm, compared with a sentence 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 (Goldberg, 2003). 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 & 002). 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 fabric of experience is. 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?

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 Y. Aloimonos (Ed.), *Active perception* (pp. 245-282). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 177-210.

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.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*(4), 767-811.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing. Where Language, Culture, and Cognition Meet* (pp. 243-268). Hillsdale, NJ: Lawrence Erlbaum Associates.

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 parser. 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.

Dikker, S., Rabagliati, H., & Pylkkanen, L. (in press). Cognition.

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 M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 111-138). Mahwah, NJ: Lawrence Erlbaum Associates.

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. A. (2002). The lexicon and the laundromat. In P. Merlo & S. Stevenson (Eds.), *The Lexical Basis of Sentence Processing* (pp. 75-94): John Benjamins.

Fodor, J. A., & Pylyshyn, Z. W. (1988). *Connectionism and cognitive architecture: A critical analysis*. Cambridge, MA: MIT Press/Bradford Books.

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.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

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 E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 173-190). 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.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: MIT Press.

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

Griffiths, T. L., & Steyvers, M. (2004). *A probabilistic approach to semantic representation.* Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society, George Mason University.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research, 16*(1), 38-50.

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

Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (in press). The wind chilled the spectators but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *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 (Ed.), *Attention and performance 12: The psychology of reading* (pp. 587-599). Hove, England UK: Lawrence Erlbaum Associates.

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 Research, 1146*, 2-22.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*: 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*. Palo Alto, CA: CSLI Publications, University.

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.

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition, 65*(2), 145-168.

Lakoff, G. (1971). Presuppositions and relative well-formedness. In D. Steinberg & L. Jakobovitz (Eds.), *Semantics* (pp. 329-340). 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. *Memory 5*, 569-599.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

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

Levin, B., & Hovav, M. R. (2005). *Argument realization*. Cambridge: 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.

McClelland, J. L., St. John, M. F., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4*, 287-336.

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.

McNeil, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeil, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

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. (1974). *A framework for representing knowledge*: Massachusetts Institute of Technology.

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* (pp. 325-347): 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. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed*

Processing: Explorations in the Microstructure of Cognition (Vol. 2, pp. 318-362). Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1988). Schemata and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 7-57). Cambridge, MA: MIT Press.

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. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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. (1993). *A Dynamic Systems Approach to Development: Applications*. Cambridge, MA: MIT 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. (2007). *The Continuity of Mind*. Oxford: Oxford University 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. (1992). The Story Gestalt: A Model of Knowledge-Intensive Processes in Text Comprehension. *Cognitive Science, 16*, 271-306.

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. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory & Language, 50*, 355-370.

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.

Tabor, W., & Tanenhaus, M. K. (2001). Dynamical systems for sentence processing. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 177-211). Westport, CT: Ablex Publishing.

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. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 414-419). Indiana University: Lawrence Erlbaum Associates.

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.