Lexical knowledge without a lexicon?

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

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.

Keywords

connectionist models; neural networks; dynamical systems; prediction; event representations; schema; sentence processing

Introduction

Words have had a checkered past, at least as objects for scientific study. For many decades, the study of words — their history, meaning, usage, etc. — constituted a main focus of linguistic research. This changed radically in the middle of the last century, starting with the publication of Chomsky's *Syntactic Structures* and *Aspects of a Theory of Syntax* (1957, 1965). Generative theories redirected the attention of linguists and psycholinguists to syntax (and to a lesser extent, semantics). Rules were where the action was, because they seemed to best account for the productive and generative nature of linguistic knowledge. Words, on the other hand, were idiosyncratic (insofar as the mapping between meaning and phonological form was mostly arbitrary and variable across languages). They had to be learned, to be sure, but their seemingly unsystematic character suggested that learning had to be rote. The mental lexicon was a rather uninteresting place, necessary but rather dull.

In recent decades, words have made a comeback. 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 fields of psycholinguistics and computational linguistics, 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, the statistical patterns of usage, and that these interactions may occur at early stages of processing (Altmann, 1998; Hare, McRae, & Elman, 2003; MacDonald, 1997; Roland & Jurafsky, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

All of this has led to a sea change, resulting in the view of the mental lexicon as a data structure of tremendous richness and detail. And this, I want to propose, should begin to raise some worries. How much detail ought to go in the lexicon? Is there a principled way to adjudicate between information that belongs in the lexicon and information that belongs elsewhere?

In the remainder of this paper, I want to suggest that the metaphor of the mental lexicon as a dictionary-like data structure is not up to the job. But before going any further, I need to make something very clear. There is no question that lexical knowledge, that is, knowledge of words' properties and appropriate usage, is extremely rich. To argue against the existence of a mental lexicon, which is what I will do, is not to argue that people lack knowledge of words. As a convenient way to talk about lexical knowledge, using the term 'mental lexicon' is not problematic. The issue I am concerned with is rather 'What is the cognitive mechanism that encodes and deploys word knowledge?' Is this knowledge really encoded in an enumerated declarative data structure, akin to a dictionary? If not, how then might lexical knowledge be instantiated? I will argue that the dictionary metaphor is ultimately infeasible, and will suggest an alternative.

What information might go in the mental lexicon?

The metaphor of the mental lexicon as a dictionary is pervasive and compelling. For many theories it not just a metaphor, but is taken seriously as a description of the data structure that stores knowledge about words. I take Jackendoff's (2002) description of the lexicon as a reasonable and typical account:

> For a first approximation, the lexicon is the store of words in long-term memory from which the grammar constructs phrases and sentences (p. 130)…[A lexical entry] lists a small chunk of phonology, a small chunk of syntax, and a small chunk of semantics (p. 131).

But just how small is the "small chunk" of phonology, syntax, and semantics? This is a crucial question. As already observed, the richness of lexical entries has grown considerably over past decades. Is there a principled limit? And are there empirical tests we might impose that would help adjudicate which information is likely to reside in a lexical entry, and which information might arise from other knowledge sources, including semantics, pragmatics, and general world knowledge? Let me turn now to knowledge of a specific lexical class of items, verbs, and discuss what recent experimental research might contribute to this discussion. I will begin by reviewing some of the findings in the sentence processing literature — particularly results dealing with verb behavior — because those have been especially significant in promoting the view of an enriched lexicon.

Sentence processing and the lexicon — background

Within the psycholinguistic literature, much of the data that led to new views about the lexicon have resulted not from the direct study of lexical representations per se, but as a by-product of a theoretical debate in recent decades 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, this leads to points within the sentence that are 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.

One hypothesis has been that processing occurs in at least two stages (e.g., Frazier, 1978, 1990, 1995; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). 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, thematic roles, etc. (Chomsky, 1965, 1981; Dowty, 1991; Katz & Fodor, 1963). Then, at a slightly later point in time, a second stage of processing occurs in which more complete 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 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.

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) 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 better model of human sentence processing. It is this research that has supported the enriched lexicon hypothesis. In what follows, I begin by describing several studies in which the results imply that 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.

Arguments for an enriched lexicon

The relationship between meaning and complement structure preferences

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, <u>*the story*</u> (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 an 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 et al., 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, in press) 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 et al. (2005).

A similar demonstration of the use of meaning to predict structure is reported in Hare et al. (in press). 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.

Verb specific thematic role filler preferences

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...* 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 et al. (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 semantic 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. McRae et al. interpreted these role-filler preferences as reflecting comprehenders' specific knowledge of the event structure associated with different verbs. This appeal to event structure, as we shall see below, will figure significantly in phenomena that are not as easily accommodated by the lexicon.

We have seen that verb-specific preferences for their thematic role fillers arise in the course of sentence processing. Might such preferences also be revealed in word-word priming? The 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, that is, 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 are but a few of very many similar experiments that have suggested that the lexical representation for verbs must include subentries about all the verb's senses. Furthermore, for each sense, all possible subcategorization frames would be shown. For each verb-sense-subcategorization combination, additional information would be indicating the probability of each combination. Finally, similar information would be needed for every

verb-sense-thematic role possibility. The experimental evidence indicates that in many cases, this latter information will be detailed, highly idiosyncratic of the verb, and represented at the featural level (e.g., Ferretti et al., 2001; McRae et al., 1997).

Challenges for the mental lexicon

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. We now turn to additional phenomena that will be problematic for the traditional view of the mental lexicon as an enumerative data structure, akin to a dictionary.

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 provide background information prefatory to the time period under focus (as in *Dorothy had skated for many years and was now looking forward to her retirement*). Imperfective aspect, on the other hand, is used to describe events that are either habitual or 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 because of the location's greater salience to the unfolding event.

This prediction was borne out. The two word prime *had skated* failed to yield significant priming for *arena* in a short SOA naming task, relative to an unrelated prime; but the two word prime *was skating* did significantly facilitate naming. In an ERP version of the experiment, the typicality of the location was found to affect expectations. Sentences such as *The diver was snorkeling in the ocean* (typical location) elicited lower amplitude N400 responses at *ocean*, compared to *The diver was snorkeling in the pond* at *pond*. The N400 is interpreted as an index of semantic expectancy, and the fact that typicality of agent-verb-location combinations affected processing at the location indicates that this information must be available early in processing.

The ability of verbal aspect to manipulate sentence processing by changing the focus on an event description can also be seen in the very different domain of pronoun interpretation. The question arises, How do comprehenders 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. (2008) suggested that, as was found in the Ferretti et al. (2007) study, aspect might alter this result. 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 ERP data indicated aspectual differences within 400 ms of the expected word's presentation). 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. In languages such as English, 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.

Dynamic alterations in verb argument expectations

If we think in terms of verbs as cues and events as the knowledge they target, then it should be clear that although the verb is obviously a very powerful cue, and 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 what a lumberjack cuts, plays an important role in comprehension at some point.

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

Agent dependencies

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

The fact that what patient is expected may vary as a function of specific particular agent-verb combinations is not in itself surprising. What is significant is that the effect occurs at the earliest possible moment, at the patient that immediately follows the verb. The timing of such effects has in the past often been taken as indicative of an effect's source. A common assumption has been that immediate effects reflect lexical or 'first-pass' processing, and later effects reflect the use of semantic or pragmatic information. In this study, the agent-verb combinations draw 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.

Instrument dependencies

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 et al. (in press) 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. 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.

Discourse dependencies

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 (2008) took a subset of the sentences used in the Bicknell et al. (2010) 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 context. The specifics of the situation in which the action occurs matter.

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

Lexical knowledge without a lexicon

Where does lexical knowledge reside?

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

Where things get sticky is when one also considers what seems to be the ability of dynamic factors to significantly modulate such expectations. These include the verb's grammatical aspect, the agent and instrument that are involved in the activity, and the overall discourse context. That these factors should play a role in sentence processing, at some point in time, is not itself surprising. The common assumption has been that such dynamic factors lie outside the lexicon. This is, for example, 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). Thus, what is significant about the findings above is that the influence of aspect, agent, instrument, and discourse all occur within the same time frame that has been used operationally to identify information that resides in the lexicon. This is important if we are to have some empirical basis for deciding what goes in the lexicon and what does not.

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

One option would be to abandon any hope of finding any empirical basis for determining the contents of the mental lexicon. One might simply stipulate that some classes of information reside in the lexicon and others do not. This is not a desirable solution. Note that even within the domain of theoretical linguistics, there has been 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; J. A. Fodor, 2002; Haiman, 1980; Jackendoff, 1983, 2002; Katz & Fodor, 1963; Lakoff, 1971; Langacker, 1987; Levin & Hovav, 2005; Weinreich, 1962). If we insist that the form of the mental lexicon has no consequences for processing, and exclude data of this type, this puts us in the awkward position where we have no behavioral way to evaluate different proposals.

A second option would be to significantly enlarge the format of lexical entries so that they accommodate all the above information. This would be a logical conclusion to the trend that

has appeared not only in the processing literature (e.g., in addition to the studies cited above, Altmann & Kamide, 2007; Kamide, Altmann, et al., 2003; Kamide, Scheepers, & Altmann, 2003; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 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?

One problem is that the combinatoric explosion this entails would be significant. In fact, given the unbounded nature of discourse contexts, it is unclear that this is even feasible. But it also presents us with a logical conundrum: If all this information resides in the lexicon, is there then any meaningful distinction between the lexicon and other linguistic modules?

The third option is the most radical. It is to consider the possibility that lexical knowledge might be instantiated in a very different way than through a mental dictionary, and to find a computational mechanism that permits the sorts of complex interactions that appear to be required to use words appropriately.

An alternative to the mental lexicon as dictionary

The common factor in the studies described above was the ability of sentential elements to interact in real time to produce an incremental interpretation that guided expectancies about upcoming elements. These can be thought of as very powerful context effects that modulate the meaning that words have. Alternatively (but equivalently) one can view words not as elements in a data structure that must be retrieved from memory, but rather as stimuli that alter mental states (which arise from processing prior words) in lawful ways. In this view, words are not mental objects that reside in a mental lexicon. They are operators on mental states.

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

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

One very simple architecture that illustrates how one might model the context- dependent nature of lexical knowledge is a connectionist model known as a *simple recurrent network* (SRN; Elman, 1990), shown schematically in Figure 1. Each rectangular box stands for some set of processing units (akin to abstract, highly simplified neurons in a real neural network) that connects to the units in other layers (shown as other rectangles). The arrows illustrate the flow of activation. What I have called the "mental state" of the system is the activation pattern that is present in the Hidden Layer at any given point in time. This internal state varies as a function of both its own prior internal state (this is the result of the feedback connections that allow the state at time t to feed into the state at time $t+1$) and the external input. In this model, the inputs correspond to words, which might be represented either abstractly as binary valued vectors but could also be represented as phonological forms. Finally, at each point in time, the network produces some output on its Output Layer.

Knowledge in such a network is contained in the pattern of connections between processing units, and in particular, in the strength of the weights. Prior to learning, the SRN's weights are initialized with small random values. At this point in time, it knows nothing. Learning is done by example. Rather than instructing the network on explicit rules, the network is shown examples of well formed stimuli (in this case, grammatical sentences). In this example, the network was presented with a large number of sentences that exemplify the ways in which verb arguments may depend on complex interactions between each other.

One task that is deceptively simple but turns out to be very powerful is prediction. In this task, the network is presented with the words in a sentence in succession. At every time step, the network is asked to predict what the next word will be. Words are represented as arbitrary binary vectors, which deprives the network of prior information regarding the lexicosemantic properties of words. A simple learning algorithm (Rumelhart, Hinton, & Williams, 1986) is used to gradually adjust the connection weights so that, over time, the network's actual output more closely appropriates the desired output (in this case, the correct next word).

If the training data are sufficiently large and complex, the network will typically not be able to memorize the sentences. Given the nondeterministic nature of most sentences, this means that the network will not be able to literally predict successive words. What we really would hope for, and what the network succeeds in doing, is to learn the context-contingent dependencies that make some words more probable successors than others, and rules out some words as ungrammatical.

For example, after learning, and given the test sentence *The girl ate the…*, the network will not predict a single word, but all possible words that are sensible in this context, given the language sample it has experienced. Thus, it might predict *sandwich*, *taco*, *cookie*, and other edible nominals. Words that are either ungrammatical (e.g., verbs) or semantically or pragmatically inappropriate (e.g., *rock*) will not be predicted.

From this behavior, we might infer the network learns the lexicogrammatical categories implicit in the training data. We can verify this by analyzing the internal representations that the network has learned for each word. These are instantiated in the activation patterns at the Hidden Layer level that are produced as the network incrementally processes successive words in sentences. These activations patterns are vectors, have a geometric interpretation as points in a high dimensional "mental space" of the network. The patterns have a similarity structure which corresponds to each word's proximity to every other word in that space.

Figure 2 displays a hierarchical clustering tree that depicts that similarity structure. Words whose internal representations are close in the "mental space" are shown as leaves that are close on the tree. One can see through the structure of the tree that the network has grouped nouns apart from verbs, and that also makes finer grained distinctions within these two categories (e.g., *animate* vs. *inanimate* nouns; *large animals* vs. *small animals; transitive* vs. *intransitive verbs*, etc.). Figure 3 shows, in simplified cartoon form and in just three dimensions, what the actual spatial relations between the words' internal representations might look like.

Not visible in either of these representations is the trajectory or path over time through the network's internal space that results when successive words in a sentence are presented. These trajectories reflect the intrinsic dynamics of the network, such that only some paths through state space are felicitous and well formed. Indeed, the network dynamics encode what we would conventionally think of as the grammar that underlies the language sample. An important discovery in recent years is that networks of this sort can implement recursive relationships that allow them to represent abstract long distance dependencies, and that the

grammars that are learned generalize beyond the training data (Boden & Blair, 2003; Boden & Wiles, 2000; Rodriguez, 2001; Rodriguez & Elman, 1999; Rodriguez, Wiles, & Elman, 1999). We will shortly see that such trajectories also play a role in encoding lexical knowledge.

Let us now consider a simple model that is exposed to language data in which knowledge of the proper use of a verb involves learning the complex dependencies between the specific arguments and adjuncts that may be used with the verb in different situations. Figure 4 is a schematic depiction of a family of sentences that illustrate such possibilities for the verb *cut*.

After training, the network is then tested by inputting, a word at a time, various sentences that illustrate these complexities. The network's predictions closely accord with what is appropriate, such that after processing *The butcher uses a saw to cut...* the network predicts that the next word will be *meat*, whereas after *A person uses a saw to cut...* the response is *a tree.* At a behavioral level, then, the network demonstrates that it has learned the lexical and grammatical regularities underlying the sentences.

How does the network do this? There are two interdependent strategies. First, the network learns to partition its internal representational space so that the internal representations that arise in real time as the network processes words reflect the basic lexico-semantic properties of the vocabulary (in much the same way as shown in Figure 3). Second, the syntagmatic knowledge governing argumentadjunct- verb interactions arise from the dynamical properties of the network (encoded in the weights between units). This means that when a word is processed, that word's impact on the internal state combines with the prior context to generate predictions about the class of grammatically appropriate continuations would be. We can see this by plotting the trajectories over time for various sentences involving the same verb but different arguments or adjuncts.

Figure 5. shows the trajectories through the network's internal state space as it processes two sentences involving the verb *cut*. We see that the states for *cut* in the two instances are close in space, reflecting the fact that there is considerable overlap in the context specific meaning associated with these two usages. But the states are not identical, illustrating the ways in which the contingencies between a verb and the other elements in a sentence combine to determine (in this case) the likely filler of the theme role. This interaction between context and lexical knowledge is inextricable and immediate.

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). Similar points have been made by many others, notably including Wittgenstein (1966), Hutchins (1994) and Fauconnier (1997; Fauconnier & Turner, 2002).

2. Context is always with us. The meaning of a word is never "out of context", although we might not always know what the context is (particularly if we fail to provide one). This point has been made by many, including Kintsch (1988), Langacker (1987), McClelland, St. John, & Taraban (1989), and van Berkum et al.

(2003; 2005). This insight is of course also what underlies computational models of meaning that emphasize multiple co-occurrence constraints between words in order to represent them as points in a high dimensional space, such as LSA (Landauer & Dumais, 1997), HAL (Burgess & Lund, 1997), or probabilistic models (Griffiths & Steyvers, 2004). The dynamical approach here also emphasizes the time course of processing that results from the incremental nature of language input.

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. These points have been made elsewhere by many, including Elman (1990), Kahneman and Tversky (1973), Kveraga, Ghuman, and Bar (2007), Schultz, Dayan, and Montague (1997), and Spirtes, Glymour, and Scheines (2000). It should not be surprising that prediction would also be exploited for language learning and play a role in on-line language comprehension.

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. An extensive literature argues for this, aside from the studies described here, including work by Minsky (1974), Schank and Abelson (1977), and Zacks and Tversky (2001); see also Shipley and Zacks (2008) for a comprehensive collection on the role of event knowledge in perception, action, and cognition.

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. Dynamical analyses figure prominently in the recent literature in cognitive science, including work by Smith and Thelen (1993, 2003; Thelen & Smith, 1994), Spencer and Schöner (2003), Spivey (2007), Spivey and Dale (2004), Tabor (2004), Tabor, Juliano, and Tanenhaus (1997) and Tabor and Tanenhaus (2001).

It must be emphasized that the model in Figure 6 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. Many important details are omitted. 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 (Clark, 1996, 2003; Goldin-Meadow, 2003; McNeil, 1992, 2005). The dynamics of such a system would be considerably more complex than those shown in Figure 6, 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.

How does this view affect the way we do business (or at least, study words)? Although I have argued that much of the behavioral phenomena described above are not easily incorporated into a mental lexicon as traditionally conceived, 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. The important question would still 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 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 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, 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 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 (Goldberg, 2003). And of course, the discourse context may provide further constraints on how an event is construed. Thus, as Race et al. (2008) found, although *shoppers* might typically save money and *lifeguards* save children, in the context of a disaster, both agents will be expected to save children.

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 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 also 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.

Acknowledgments

References

Altmann GTM. Ambiguity in sentence processing. Trends in Cognitive Sciences. 1998; 2:146–152. [PubMed: 21227111]

Altmann GTM. Thematic role assignment in context. Journal of Memory & Language. 1999; 41(1): 124–145.

Altmann GTM, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition. 1999; 73(3):247–264. [PubMed: 10585516]

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

Ambridge B, Pine JM, Rowland CF, Young CR. The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure. Cognition. 2008; 106(1):87–129. [PubMed: 17316595]

Anderson, SJ.; Conway, MA. Representations of autobiographical memories. In: Conway, MA., editor. Cognitive Models of Memory. Cambridge, MA: MIT Press; 1997. p. 217-246.

Ballard, DH.; Brown, CM. Principles of animate vision. In: Aloimonos, Y., editor. Active perception. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 245-282.

Bates E, Goodman JC. On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. Language & Cognitive Processes. 1997; 12:507–584.

Bicknell K, Elman JL, Hare M, McRae K, Kutas M. Effects of event knowledge in processing verbal arguments. Journal of Memory & Language. 2010; 63(4):489–505. [PubMed: 21076629]

Boden M, Blair A. Learning the dynamics of embedded clauses. Applied Intelligence. 2003; 19(1–2): 51–63.

Boden M, Wiles J. Context-free and context-sensitive dynamics in recurrent neural networks. Connection Science: Journal of Neural Computing, Artificial Intelligence & Cognitive Research. 2000; 12(3–4):197–210.

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

Brown NR, Schopflocher D. Event clusters: An organization of personal events in autobiographical memory. Psychological Science. 1998; 9:470–475.

Burgess C, Lund K. Modeling parsing constraints with high-dimensional context space. Language & Cognitive Processes. 1997; 12:177–210.

Chomsky, N. Syntactic structures. The Hague: Mouton; 1957.

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

Chomsky, N. Rules and Representations. New York: Columbia University Press; 1980.

Chomsky, N. Lectures on Government and Binding. New York: Foris; 1981.

Clark, HH. Using Language. Cambridge: Cambridge University Press; 1996.

Clark, HH. Pointing and placing. In: Kita, S., editor. Pointing. Where Language, Culture, and Cognition Meet. Hillsdale, NJ: Lawrence Erlbaum Associates; 2003. p. 243-268.

Crain S. Language acquisition in the absence of experience. Brain & Behavioral Sciences. 1991; 14:597–611.

Crain, S.; Steedman, M. On not being led up the garden path: The use of context by the psychological parser. In: Dowty, D.; Karttunen, L.; Zwicky, A., editors. Natural Language Processing: Psychological, Computational, and Theoretical Perspectives. Cambridge: Cambridge University Press; 1985. p. 320-358.

de Sa, VR. Sensory modality segregation. In: Thurn, S.; Saul, L.; Schoelkopf, B., editors. Advances in Neural Information Processing Systems. Vol. 16. Cambridge, MA: MIT Press; 2004. p. 913-920.

de Sa VR, Ballard DH. Category learning through multimodality sensing. Neural Computation. 1998; 10(5):1097–1117. [PubMed: 9654768]

Dowty D. Thematic proto-roles and argument selection. Language. 1991; 67:547–619.

Elman JL. Finding Structure in Time. Cognitive Science. 1990; 14(2):179–211.

Elman, JL.; Hare, M.; McRae, K. Cues, constraints, and competition in sentence processing. In: Tomasello, M.; Slobin, D., editors. Beyond nature-nurture: Essays in honor of Elizabeth Bates. Mahwah, NJ: Lawrence Erlbaum Associates; 2005. p. 111-138.

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

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

Ferreira F, Henderson JM. 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. 1990; 16(4):555–568.

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

Ferretti TR, McRae K, Hatherell A. Integrating verbs, situation schemas, and thematic role concepts. Journal of Memory & Language. 2001; 44(4):516–547.

Fodor, JA. The lexicon and the laundromat. In: Merlo, P.; Stevenson, S., editors. The Lexical Basis of Sentence Processing. Amsterdam: John Benjamins; 2002. p. 75-94.

Fodor, JD. Thematic roles and modularity. In: Altmann, GTM., editor. Cognitive Models of Speech Processing. Cambridge, MA: MIT Press; 1995. p. 434-456.

Ford, M.; Bresnan, J.; Kaplan, RM. A competence-based theory of syntactic closure. In: Bresnan, J., editor. The Mental Representation of Grammatical Relations. Cambridge: MIT Press; 1982. p. 727-796.

Frazier, L. Unpublished PhD. University of Connecticut; 1978. On comprehending sentences: Syntactic parsing strategies.

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

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

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

Garnsey SM, Pearlmutter NJ, Meyers E, Lotocky MA. The contribution of verbbias and plausibility to the comprehension of temporarily ambiguous sentences. Journal of Memory & Language. 1997; 37:58–93.

Gold EM. Language identification in the limit. Information & Control. 1967; 16:447–474.

Goldberg AE. Constructions: a new theoretical approach to language. Trends in Cognitive Sciences. 2003; 7(5):219–224. [PubMed: 12757824]

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

Griffiths, TL.; Steyvers, M. A probabilistic approach to semantic representation. Paper presented at the Proceedings of the 24th Annual Conference of the Cognitive Science Society; George Mason University; 2004.

Haiman J. Dictionaries and encyclopedias. Lingua. 1980; 50:329–357.

Hare M, Elman JL, Tabaczynski T, McRae K. The wind chilled the spectators but the wine just chilled: Sense, structure, and sentence comprehension. Cognitive Science. in press.

Hare M, McRae K, Elman JL. Sense and structure: Meaning as a determinant of verb subcategorization preferences. Journal of Memory & Language. 2003; 48(2):281–303.

Hare M, McRae K, Elman JL. Admitting that admitting verb sense into corpus analyses makes sense. Language & Cognitive Processes. 2004; 19:181–224.

Hobbs, JR. Literature and cognition. Stanford, CA: Center for the Study of Language and Information; 1990.

Holmes, VM. Syntactic parsing: In search of the garden path. In: Coltheart, M., editor. Attention and performance 12: The psychology of reading. Hove, England UK: Lawrence Erlbaum Associates; 1987. p. 587-599.

Hutchins, E. Cognition in the World. Cambridge, MA: MIT Press; 1994.

Jackendoff, R. Semantics and cognition. Cambridge, MA: MIT Press; 1983.

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

Jackendoff R. A parallel architecture perspective on language processing. Brain Research. 2007; 1146:2–22. [PubMed: 17045978]

Kahneman D, Tversky A. On the psychology of prediction. Psychological Review. 1973; 80(4):237–251.

Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory & Language. 2003; 49(1):133–156.

Kamide Y, Scheepers C, Altmann GTM. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. Journal of Psycholinguistic Research. 2003; 32(1):37–55. [PubMed: 12647562]

Katz JJ, Fodor JA. The structure of a semantic theory. Language. 1963; 39(2):170–210.

Kehler, A. Coherence, reference, and the theory of grammar. Palo Alto, CA: CSLI Publications, University; 2002.

Kehler A, Kertz L, Rohde H, Elman JL. Coherence and coreference revisited. Journal of Semantics. 2008; 25:1–44.

Kennison SM. American English usage frequencies for noun phrase and tensed sentence complement-taking verbs. Journal of Psycholinguistic Research. 1999; 28(2):165–177. [PubMed: 10474887]

Kintsch W. The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review. 1988; 95(2):163–182. [PubMed: 3375398]

Kveraga K, Ghuman AS, Bar M. Top-down predictions in the cognitive brain. Brain & Cognition. 2007; 65(2):145–168. [PubMed: 17923222]

Lakoff, G. Presuppositions and relative well-formedness. In: Steinberg, D.; Jakobovitz, L., editors. Semantics. London: Cambridge University Press; 1971. p. 329-340.

Lakoff, G. Women, fire, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press; 1987.

Lancaster JS, Barsalou L. Multiple organisations of events in memory. Memory. 1997; 5:569–599. [PubMed: 9415322]

Landauer TK, Dumais ST. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review. 1997; 104:211–240.

Langacker, RW. Foundations of Cognitive Grammar. Vol. 1. Stanford: Stanford University Press; 1987.

Levin, B.; Hovav, MR. Argument realization. Cambridge: Cambridge University Press; 2005.

Lewis, JD.; Elman, JL. A connectionist investigation of linguistic arguments from the poverty of the stimulus: Learning the unlearnable. In: Moore, JD.; Stenning, K., editors. Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 552-557.

MacDonald MC. The interaction of lexical and syntactic ambiguity. Journal of Memory & Language. 1993; 32(5):692–715.

MacDonald MC. Lexical representations and sentence processing: An introduction. Language & Cognitive Processes. 1997; 12(2&3):121–136.

MacDonald MC, Pearlmutter NJ, Seidenberg MS. The lexical nature of syntactic ambiguity resolution. Psychological Review. 1994; 101(4):676–703. [PubMed: 7984711]

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

Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Dynamic combination of concepts during on-line language comprehension. Journal of Experimental Psychology: Learning, Memory and Cognition. in press.

McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review. 1981; 86:287–330.

McClelland JL, St John MF, Taraban R. Sentence comprehension: A parallel distributed processing approach. Language & Cognitive Processes. 1989; 4:287–336.

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

McNeil, D. Gesture and thought. Chicago: University of Chicago Press; 2005.

McRae K, Ferretti TR, Amyote L. Thematic roles as verb-specific concepts. Language and Cognitive Processes: Special Issue on Lexical Representations in Sentence Processing. 1997; 12:137–176.

McRae K, Hare M, Elman JL, Ferretti TR. A basis for generating expectancies for verbs from nouns. Memory & Cognition. 2005; 33(7):1174–1184.

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

Minsky, M. A framework for representing knowledge. Massachusetts Institute of Technology; 1974.

Mitchell, DC. Lexical guidance in human parsing: Locus and processing characteristics. In: Coltheart, M., editor. Attention and Performance XII: The psychology of reading. Hillsdale, NJ: Erlbaum; 1987. p. 601-618.

Mitchell DC, Holmes VM. The role of specific information about the verb in parsing sentences with local ambiguity. Journal of Memory & Language. 1985; 24:542–559.

Pickering MJ, Traxler MJ. Plausibility and recovery from garden-paths: An eyetracking study. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1998; 24:940–961.

Pullum GK, Scholz BC. Empirical assessment of stimulus poverty arguments. Linguistic Review. 2002; 19(1–2):9–50.

Race, D.; Klein, N.; Hare, M.; Tanenhaus, MK. What do shoppers expect to save? Agents influence patients via events, not associative priming. Poster presented at the 20th Annual CUNY Conference on Human Sentence Processing; Chapel Hill, NC. 2008.

Rayner K, Carlson M, Frazier L. The interaction of syntax and semantics during sentence processing. Journal of Verbal Learning & Verbal Behavior. 1983; 22:358–374.

Reali F, Christiansen MH. Uncovering the Richness of the Stimulus: Structure Dependence and Indirect Statistical Evidence. Cognitive Science: A Multidisciplinary Journal. 2005; 29(6):1007–1028.

Reiser BJ, Black JB, Abelson RP. Knowledge structures in the organization and retrieval of autobiographical memory. Cognitive Psychology. 1985; 17:89–137.

Rodriguez P. Simple recurrent networks learn context-free and context-sensitive languages by counting. Neural Computation. 2001; 13(9):2093–2118. [PubMed: 11516359]

Rodriguez P, Elman JL. Watching the transients: viewing a simple recurrent network as a limited counter. Behaviormetrika. 1999; 26(1):51–74.

Rodriguez P, Wiles J, Elman JL. A recurrent neural network that learns to count. Connection Science. 1999; 11(1):5–40.

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

Roland, D.; Jurafsky, D. Verb sense and verb subcategorization probabilities. In: Merlo, P.; Stevenson, S., editors. The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues. Amsterdam: John Benjamins; 2002. p. 325-347.

Rumelhart, DE. Some problems with the notion that words have literal meanings. In: Ortony, A., editor. Metaphor and Thought. Cambridge: Cambridge University Press; 1979. p. 71-82.

Rumelhart, DE.; Hinton, GE.; Williams, RJ. Learning internal representations by error propagation. In: Rumelhart, DE.; McClelland, JL., editors. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2. Cambridge, MA: MIT Press; 1986. p. 318-362.

Rumelhart, DE.; Smolensky, P.; McClelland, JL.; Hinton, GE. Schemata and sequential thought processes in PDP models. In: McClelland, JL.; Rumelhart, DE., editors. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2. Cambridge, MA: MIT Press; 1988. p. 7-57.

Schank, RC.; Abelson, RP. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977.

Schmauder AR, Egan MC. The influence of semantic fit on on-line sentence processing. Memory & Cognition. 1998; 26(6):1304–1312.

Scholz BC, Pullum GK. Searching for arguments to support linguistic nativism. Linguistic Review. 2002; 19(1–2):185–223.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275(5306):1593–1599. [PubMed: 9054347]

Shipley, TF.; Zacks, JM., editors. Understanding Events: How Humans See, Represent, and Act on Events. Oxford: Oxford University Press; 2008.

Smith, LB.; Thelen, E. A Dynamic Systems Approach to Development: Applications. Cambridge, MA: MIT Press; 1993.

Smith LB, Thelen E. Development as a dynamic system. Trends in Cognitive Sciences. 2003; 7(8): 343–348. [PubMed: 12907229]

Spencer JP, Schöner G. Bridging the representational gap in the dynamic systems approach to development. Developmental Science. 2003; 6(4):392–412.

Spirtes, P.; Glymour, CN.; Scheines, R. Causation, Prediction, and Search. Cambridge, MA: MIT Press; 2000.

Spivey, MJ. The Continuity of Mind. Oxford: Oxford University Press; 2007.

Spivey, MJ.; Dale, R., editors. On the continuity of mind: Toward a dynamical account of cognition. San Diego, CA: Elsevier Academic Press; 2004.

St John M, McClelland JL. Learning and applying contextual constraints in sentence comprehension. Artificial Intelligence. 1990; 46:217–257.

Stevenson RJ, Crawley RA, Kleinman D. Thematic roles, focus and the representation of events. Language & Cognitive Processes. 1994; 9(4):519–548.

Tabor W. Effects of merely local syntactic coherence on sentence processing. Journal of Memory & Language. 2004; 50:355–370.

Tabor W, Juliano C, Tanenhaus MK. Parsing in a dynamical system: An attractorbased account of the interaction of lexical and structural constraints in sentence processing. Language & Cognitive Processes. 1997; 12(2&3):211–271.

Tabor, W.; Tanenhaus, MK. Dynamical systems for sentence processing. In: Christiansen, MH.; Chater, N., editors. Connectionist psycholinguistics. Westport, CT: Ablex Publishing; 2001. p. 177-211.

Tanenhaus, MK.; Carlson, G. Lexical structure and language comprehension. Cambridge, MA: MIT Press; 1989.

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268(5217):1632–1634. [PubMed: 7777863]

Taraban R, McClelland JL. Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. Journal of Memory & Language. 1988; 27(6):597–632.

Thelen, E.; Smith, LB. A dynamic Systems Approach to the Development of Cognition and Action. Cambridge, MA: MIT Press; 1994.

Tomasello M. The item based nature of children's early syntactic development. Trends in Cognitive Science. 2000; 4:156–163.

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

Trueswell JC, Tanenhaus MK, Garnsey SM. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. Journal of Memory & Language. 1994; 33(3):285–318.

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

van Berkum JJA, Brown CM, Zwitserlood P, Kooijman V, Hagoort P. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory and Cognition. 2005; 31(3):443–467.

van Berkum JJA, Zwitserlood P, Hagoort P, Brown CM. When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. Cognitive Brain Research. 2003; 17(3): 701–718. [PubMed: 14561457]

Weinreich, U. Lexicographic definition in descriptive semantics. In: Householder, FW.; Saporta, W., editors. Problems in lexicography. Vol. 21. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics; 1962.

Wittgenstein, L. Philosophical investigations. Oxford, England: Macmillian; 1966.

Zacks JM, Tversky B. Event structure in perception and conception. Psychological Bulletin. 2001; 127(1):3–21. [PubMed: 11271755]
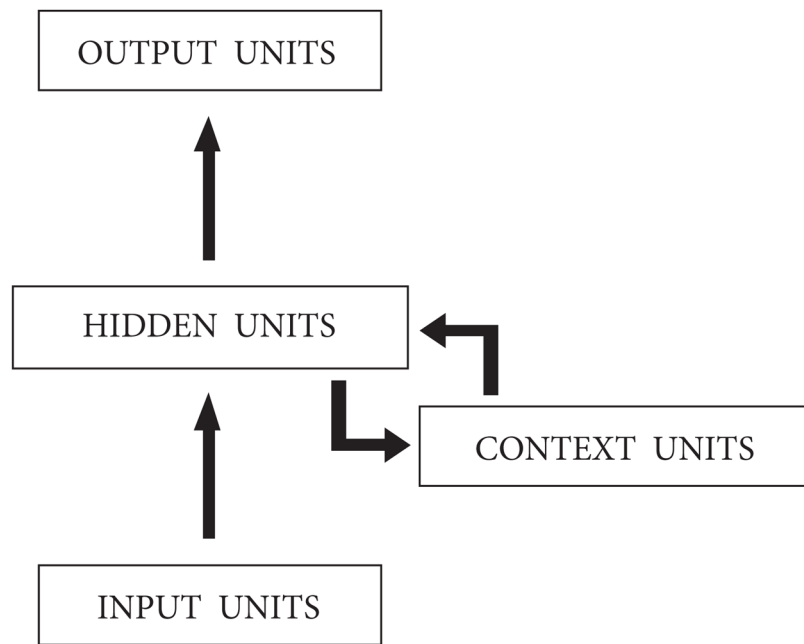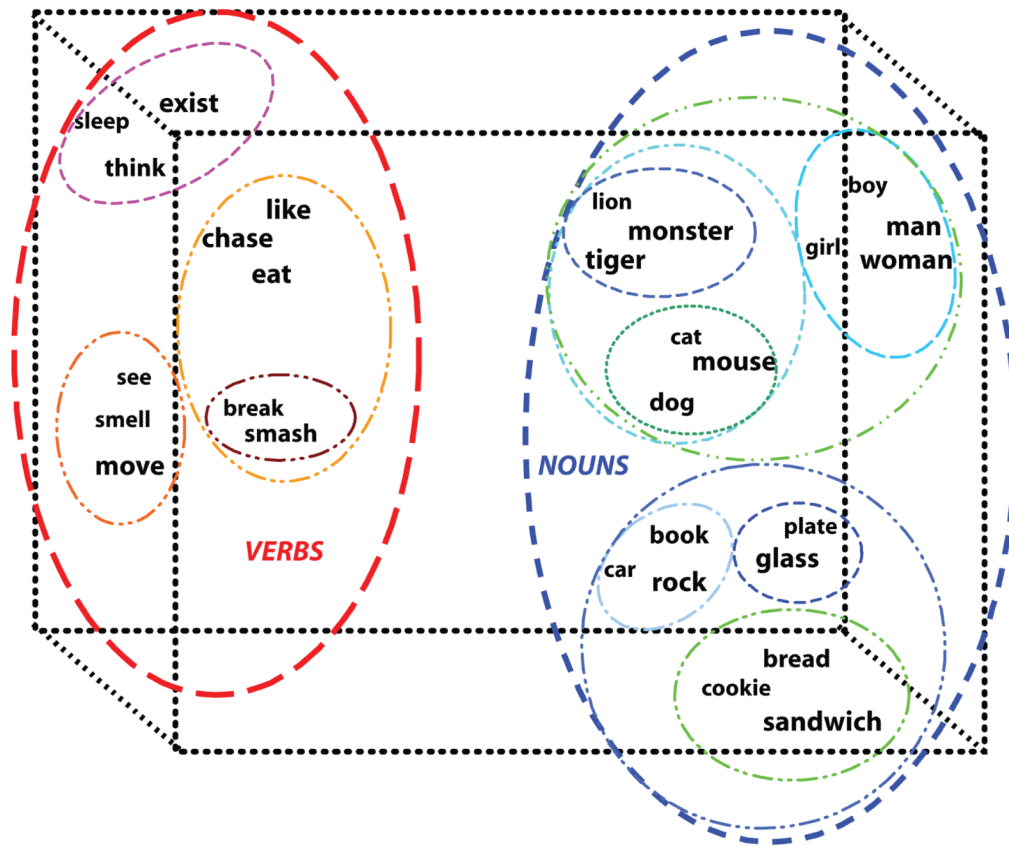
Figure 1.
Simple Recurrent Network. Each layer is composed of one or more units. Information flows from input to hidden to output layers. In addition, at every time step t, the hidden unit layer receives input from the context layer, which stores the hidden unit activations from time t – 1.
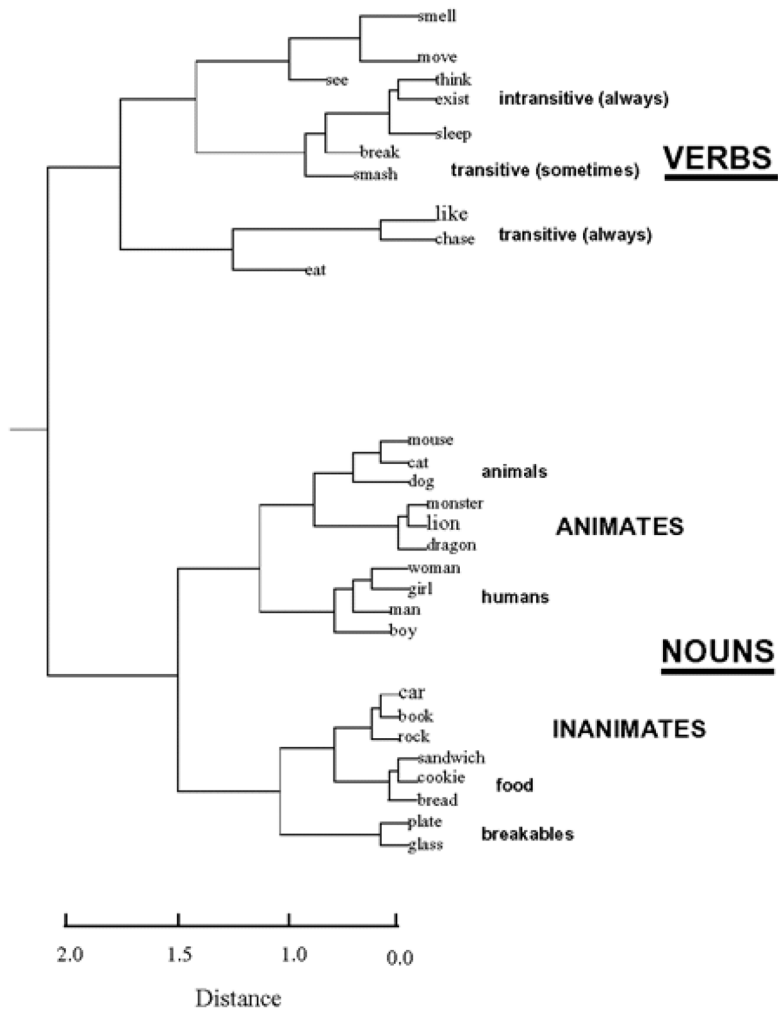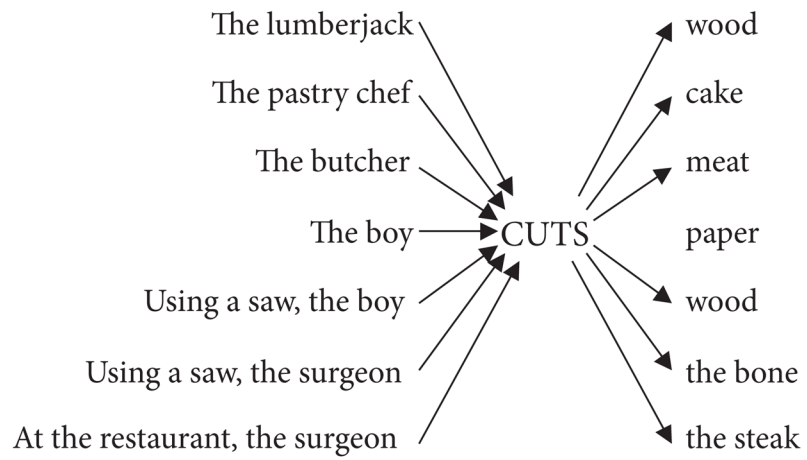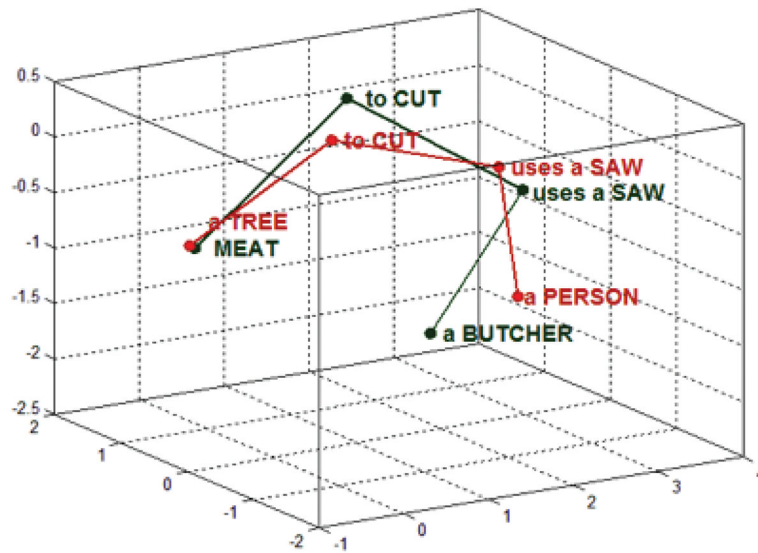
Figure 2.
Hierarchical clustering diagram of hidden unit activation patterns in response to different words. The similarity between words and groups of words is reflected in the tree structure; items that are closer are joined lower in the tree.

Figure 3.
Schematic visualization, in 3-D, of the high dimensional state space described by the SRN's hidden unit layer. The state space is partitioned into different regions that correspond to grammatical and semantic catgories. Nesting relationship in space (e.g., HUMAN within ANIMATE within NOUN categories) reflect hierarchical relationships between categories.

The lumberjack
The pastry chef
The butcher
The boy —→ CUTS
Using a saw, the boy
Using a saw, the surgeon
At the restaurant, the surgeon

wood
cake
meat
paper
wood
the bone
the steak

Figure 4.
The verb cut may denote many different types of activities, depending on (among other factors) the agent, instrument, or location with which the verb occurs. These dependencies then affect what is the likely filler of the theme role for the verb.

Figure 5.
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.