The Representation of Knowledge in Minds and Machines

Walter Kintsch
Institute of Cognitive Science, University of Colorado, Boulder, USA

Human knowledge can be represented as a propositional network in which the meaning of a node is defined by its position in the network. That is, the relationship between a node and its neighbours determines how this node is used in language understanding and production, i.e., its meaning. The propositions that make up such a network are predicate-argument structures with time and location slots. Schemas, frames, and production rules can be expressed in the same formalism. Implications for this contextual view of meaning are discussed. Since the construction of such a propositional network depends on hand coding and is therefore impractical, an alternative automatic statistical procedure is explored that yields a high-dimensional semantic space. Vectors in this space correspond to nodes in the propositional network, in that the meaning of a vector in the Latent Semantic Analysis space is given by its neighbouring vectors in that space.

Les connaissances humaines peuvent être représentées sous la forme d'un réseau propositionnel dans lequel la signification d'un noeud est définie par sa position au sein du réseau. La relation entre un noeud et ses voisins détermine la manière dont ce noeud est utilisé dans la compréhension et la production du langage, en d'autres termes, sa signification. Les propositions qui constituent un tel réseau sont des structures prédicat-argument incluant une information temporelle et une information de localisation. Les schémas, les "frames" et les règles de production sont également exprimables à l'aide de ce formalisme. Les implications de cette interprétation contextuelle de la signification sont discutées dans le présent article. Si la construction d'un tel réseau propositionnel dépendait uniquement d'un codage manuel, elle serait de fait pratiquement irréalisable. C'est pourquoi une procédure statistique automatique est envisagée. Cette procédure produit un espace sémantique à dimensions multiples. Les vecteurs dans cet espace correspondent aux noeuds du réseau propositionnel. Ainsi, la signification d'un vecteur dans l'espace de l'Analyse Sémantique Latente est donnée par les vecteurs voisins au sein de cet espace.

How is knowledge represented in the human mind? How should knowledge be represented in cognitive science theories, and in machines that simulate human knowledge? Philosophers have asked such questions for a long time and in recent years these questions have moved to centre stage in cognitive science. Linguists, computer scientists, psychologists, and other cognitive scientists have realized that these questions are of fundamental significance to their work. A multitude of answers have been proposed by different scientists in different disciplines to these questions. These cannot be reviewed here. Instead, I shall focus on two related approaches that appear to be particularly useful in research on discourse comprehension. One is derived from linguistics but has been most influential in psychology and artificial intelligence: the propositional representations of the meaning of texts as well as knowledge in general. Such representations have played a major role in recent years, especially within the community of scholars interested in language comprehension. Propositional representations involve hand-coding, however, which is a serious limitation.

Requests for reprints should be addressed to Walter Kintsch, Institute of Cognitive Science, University of Colorado at Boulder, Muenzinger Psychology Building, Campus Box 344, Boulder, Colorado 80309–0344, USA.
(E-mail: wkintsch@clipr.colorado.edu)
This paper is an excerpt of Chapter 3 from *Comprehension: A paradigm for cognition*, Cambridge University Press, 1998.

Therefore, a second related form of knowledge representation is discussed here: Latent Semantic Analysis, which is essentially statistical in nature and which overcomes this limitation.

Propositional text representations (Kintsch, 1974, 1998) are helpful for describing the semantic characteristics of experimental texts, and they are useful for scoring recall and other data from experimental subjects. Furthermore, they are widely used for modelling purposes. Although they may not be ideal for every purpose, they have proven to be a reliable workhorse. In contrast, propositional knowledge representations remain largely untried. Propositional knowledge nets seem quite appropriate for the purpose of simulating knowledge use in comprehension.

THE CONSTRUCTION OF MEANING

I propose to represent knowledge as a network of propositions. Such a network is called a *knowledge net*. The nodes of the net are propositions, schemas, frames, scripts, production rules—which can all be written in a formalism based on the predicate-argument schema. The links are unlabelled and vary in strength, that is, a knowledge net is a type of associative net.

The meaning of a node is given by its position in the net, that is by the strengths with which it is linked to its neighbours—immediate ones as well as neighbours many steps apart. This definition of meaning is an abstract, linguistic one, not a psychological one. Psychologically, only those nodes that are actually active (that is, are held in working memory) contribute to the meaning of a node. Since the capacity of working memory is severely limited, any node at any point in time has only a few neighbours; its meaning is sparse, therefore. However, it can be readily elaborated, almost without limit, in many different directions as the situation demands, because most nodes in a knowledge network are connected with powerful, stable links—retrieval structures—to other nodes in the net, which can be brought into working memory. Thus, very complex meanings can be generated automatically and effortlessly, although at any particular point in time only a few nodes can be active in working memory.

Concepts do not have a fixed and permanent meaning. Rather, each time a concept is used, its meaning is constructed in working memory by activating a certain subset of the propositions in the neighbourhood of a concept node. The context of use determines which nodes linked to a concept are activated when a concept is used. Goals, prior experience, emotional state, and situational as well as semantic context all influence which nodes are activated and hence what the meaning of the concept will be on this particular occasion.

In constructing the meaning of a concept, the concept node and any other currently active proposition in working memory serve as the retrieval cues, both individually and as parts of a compound cue. Hence what will be retrieved in the process of elaborating the meaning of a node will depend not only on the node itself but on the state of working memory as a whole.

The substructure from which the meaning of a concept is constructed—the knowledge net—is relatively permanent (experience and learning creates and continually modifies this structure). The meaning, that is the portion of the knowledge net that is activated, is flexible, changeable, and temporary, however. Since meaning construction is based on the same substructure, there will be a certain amount of consistency in the meaning of concepts on different occasions. The likelihood that certain meaning elements will be sampled will always be greater than for other elements, but the context in which this sampling occurs will ensure a great deal of variability in the outcome.

Knowledge nets have some advantages over alternative forms of knowledge representation. First of all, there are good psychological data that argue for the psychological validity of each one of the knowledge structures mentioned, from associative nets to production systems. But at the same time, there are equally good data that show that neither of these systems is sufficient by itself for the representation of knowledge. For instance, one can show, as Collins and Quillian (1969) have done, that certain psychological predictions that can be derived from semantic nets can be verified experimentally, but as an army of their detractors have demonstrated, and as we can read in every cognitive psychology textbook, these predictions are wrong in many ways, because there is much more to human knowledge than a semantic net. A knowledge system must account for the inheritance of (some) properties, but it also must include schema-like structures with default-slots and procedural knowledge that links cognition and action, and so on. In such a net one can make distinctions between episodic and semantic memory, or procedural and declarative knowledge, but it is always necessary to do so. These are merely different types of nodes in the same network, interacting with each other. Knowledge nets, therefore, freely combine features from other knowledge representation systems that were shown to be useful for computational purposes in AI research, or valid in psychological experiments.

Consider a particular node P in such a net (Fig. 1). It may be linked to a set of nodes P_1, the P_1 nodes in turn are linked to another set of nodes P_2, and so on in ever-increasing concentric circles until the whole network is included. To say that the meaning of P is the relation between P and these concentric sets of nodes is true in some abstract sense but besides the point psychologically. At any given moment in time, in any given knowledge net, only those nodes contribute to the psychological meaning of P that are actually activated in working memory and linked to it. Thus, typically, instead of a very large set of nodes, only a limited number of nodes constitute the effective meaning of P at any time, perhaps only five or six nodes in the case when the meaning of P is only superficially elaborated (Fig. 2). However, because P is embedded in a network of strong stable relationships with other nodes in the knowledge net, further elaboration via retrieval structures is readily achieved should there be a need for it. I may think of a few things concerning P and you may think of a few things; there may be little overlap and a miscommunication may result. However, if there is some context to guide us, we are more likely to construct similar meanings for P and communicate effectively.

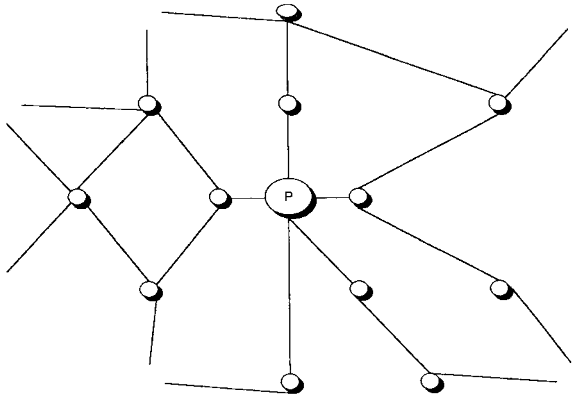

FIG. 1. The potential meaning of a proposition P is given by its position in a knowledge net.

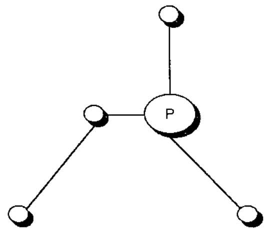

FIG. 2. The actual, momentary meaning of a proposition is the activated neighbourhood of the proposition in the knowledge net.

Knowledge nets thus imply a commitment to a radical constructionist position in the controversy about the mental representation of word meanings. In a lexicon, one looks up the meaning of a word. In a knowledge net, there is nothing to look up. Meaning has to be constructed by activating nodes in the neighbourhood of a word. This activation process is probabilistic, with activation probabilities being proportional to the strengths of connections among the nodes, and may continue for a variable amount of time and spread outwards into the knowledge net from the source node. The meaning of the source word is then the set of activated nodes in the knowledge net.

The knowledge net serves as a retrieval structure in the sense of Ericsson and Kintsch (1995). If any element of a knowledge net is in working memory (focus of attention, consciousness), other elements directly connected with it can be retrieved with a single 400msec retrieval operation. These directly retrievable elements make up what Ericsson and Kintsch called long-term working memory. Of course, once a particular element from long-term working memory has actually been instated in the focus of attention, it, too, will provide access to its neighbours in the knowledge net, thereby further increasing the size of long-term working memory. In addition, the pair of nodes will function as a compound cue, markedly changing the retrieval probabilities in long-term working memory.

Words, in this view, have a potential meaning given by concentric shells of ever-expanding neighbourhoods in the knowledge net. The most restrictive potential meaning of a concept would be given by its immediate neighbours; the most complete by the total net. There is no sense in asking how many steps away from a concept the expansion has to go to give us "the meaning" of a concept. Meaning may be more or less elaborate. And, most importantly, this expansion process provides only a potential meaning. The real, actual meaning of a word is not the set of all nodes that might be activated in long-term working memory, but rather the nodes that have actually been activated in the particular context of use. Thus, meanings are not nearly as elaborate as they could be, because normally only an insignificant fraction of a concept's neighbouring nodes in a knowledge net enters consciousness (though many more are readily available in long-term working memory). A linguist, semanticist, or psychologist studying the meaning of a concept will come up with a very rich and complex structure. That is not, however, what is actually operative when that word is used on specific occasions in a specific context, where meaning is much more sketchy and incomplete.

Contextual word meanings are not only shallow, but are dynamic and fluctuating. Somewhat different word meanings are constructed on different occasions, even if the knowledge net and discourse context remain the same, simply because of the probabilistic nature of the sampling process that determines which of the many possible knowledge elements actually enter consciousness. But the discourse context is in continuous flux, and different persons operate with different knowledge nets. Hence there must be considerable variability in effective word meanings.

Before continuing this discussion on the construction of meaning it seems only proper to ask whether there is any psychological evidence that would support such a theory. Indeed, there is. In fact, the psychological evidence overwhelmingly favours the view that concepts are temporary constructions in working memory, generated in response to task demand and subject to the constraints exercised by the underlying knowledge base and the situational context. Cognitive scientists might as well discard the traditional notion that concepts are stable entities to be retrieved from long-term memory—a view that we have inherited from philosophy and linguistics.

Barclay, Bransford, Franks, McCarrell, and Nitsch (1974) were the first to point out the role of encoding variability in memory retrieval. They gave subjects words like "*piano*" to study in the context of playing music or moving furniture. On a later memory test, they gave either "*loud*" or "*heavy*" as a retrieval cue. The former was a better retrieval cue when "*piano*" had been presented in the music context, but the latter was the better cue when "*piano*" had been studied in the furniture context, suggesting that a context-specific concept of "*piano*" was encoded.

In several studies that demonstrate the flexibility and context-dependency of concepts, a sentence verification paradigm was used. A representative experiment is that of McKoon

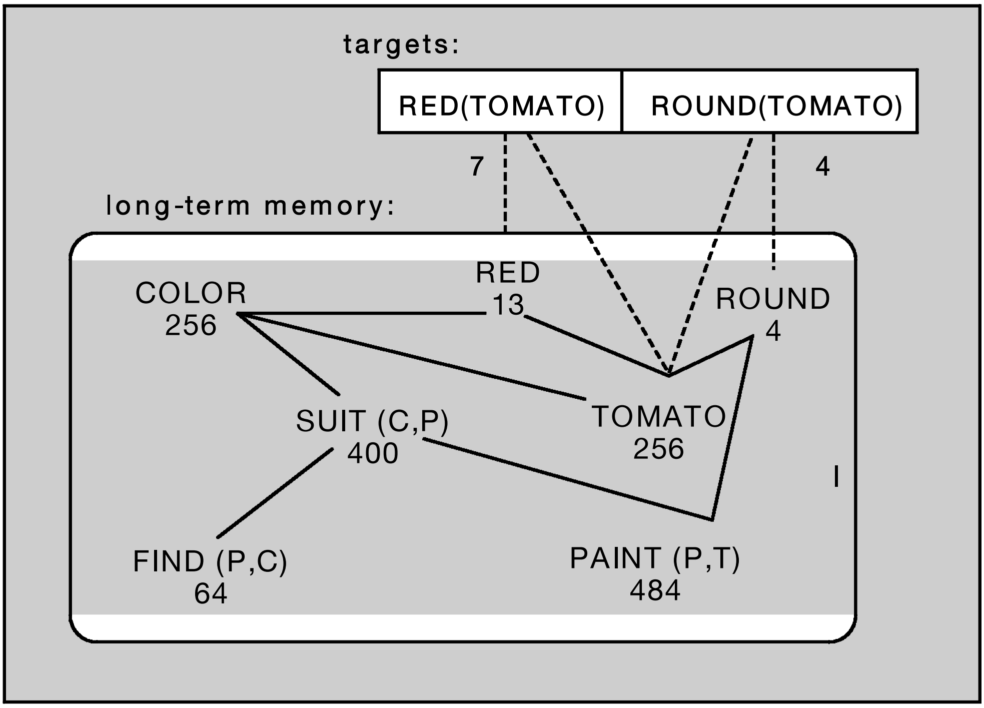

FIG. 3. Part of the long-term memory network after reading the painting-tomatoes paragraph and two target sentences. The numbers below the nodes in the network are long-term memory strengths; the numbers below the target sentences are the activation values they receive from the net.

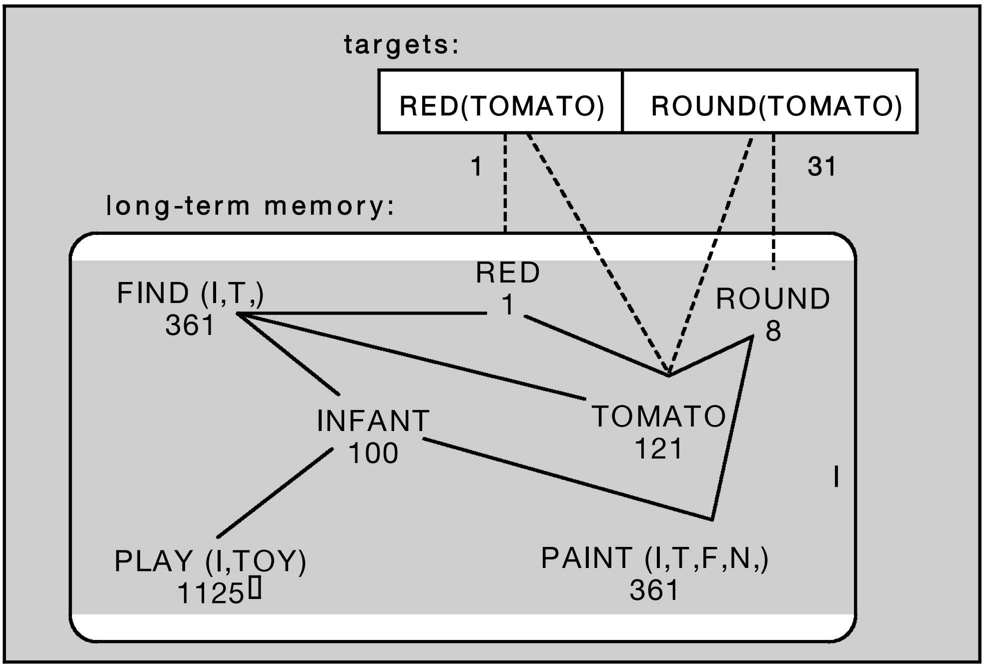

FIG. 4. Part of the long-term memory network after reading the playing-with-tomatoes paragraph and two target sentences. The numbers below the nodes in the network are long-term memory strengths; the numbers below the target sentences are the activation values they receive from the net.

and Ratcliff (1988). They argued that two equally well-known properties of tomatoes are that they are round and red. Nevertheless, the actual availability of these properties strongly depends on the context in which "*tomato*" was used. Following a brief paragraph about painting a still life containing a tomato, "*Tomatoes are red*" was verified faster than "*Tomatoes are round*". But after a paragraph describing a child rolling tomatoes around the floor, "*Tomatoes are round*" was verified faster. The average difference between matching and nonmatching sentences was 120msec. In a neutral context that emphasized neither colour nor shape (about eating tomatoes), both target sentences were responded to equally quickly.

Kintsch and Welsch (1991) have shown that these results follow directly from the assumptions about meaning construction that were discussed above. Figure 3 shows a fragment of the episodic memory trace that results from the painting-tomatoes text of McKoon and Ratcliff according to the construction-integration model. The long-term memory strength values of the propositions are also shown. More activation from this network will flow to a test sentence "*Tomatoes are red*" than to "*Tomatoes are round*". On the other hand, in the context of the playing-with-tomatoes text, the opposite will be true. The episodic memory trace for this case is shown in Fig. 4. In the context of the eating-tomatoes text (not shown here), the activation values for the two test sentences turn out to be equal.

A number of other studies demonstrating the context dependency of concepts have been reviewed by Barsalou (1993), who advocates a view related to the present one. Of particular significance is one of his own studies, which he discusses. In this experiment, subjects were asked to write down features that define common categories. Only 44% of the features in one subject's description existed in another subject's description, indicating that the definitions subjects provided were highly idiosyncratic. Indeed, when subjects were asked to provide definitions for the same concepts on two successive occasions, their own overlap was merely 66%. Thus, not only are concepts idiosyncratic, but they are also highly unstable. However, Barsalou was able to show that only the concepts that subjects constructed in this underconstrained experimental situation were so unpredictable. If subjects were given all the features generated by everyone in the experiment, they agreed very well that these were indeed features characteristic of the concept in question (97% agreement between subjects, 98% within subjects). Everyone has more or less the same knowledge base when it comes to these familiar everyday concepts. Furthermore, in a more constraining context, different subjects tended to agree much better on the features that characterize a concept. Thus, in the experiment reported by Barsalou, inter-subject agreement was 45% when subjects defined concepts like "*vehicle*" in isolation, but rose to 70% with even a minimally constraining context ("*Taking a vacation in the rugged mountains of Mexico*").

Barsalou's data nicely illustrate the fact that although concepts are fleeting and flexible, all is not chaos, because the knowledge bases from which these concepts are constructed are more stable and predictable, and most of the time the context itself will be sufficiently constraining to ensure that the concepts different people form will be similar. Nevertheless, concepts are never quite the same—surely a limiting factor in communication.

The theory of meaning advocated here is not only a constructivist, but also a minimalist, one. Clearly readers can study a text over and over again, and construct very elaborate meanings for its propositions and concepts. Linguists, philosophers, and literary critics do this all the time, and most people do so at least some of the time. But most of the time, in reading or conversation, the process of meaning construction remains shallow, not just because comprehenders are inherently lazy, but mostly because no more is required. A slight knowledge elaboration of a text is usually quite sufficient for whatever action is intended. Most of the time texts do not need much elaboration and interpretation to arrive at stable interpretations upon which appropriate responses can be based. Long-term working memory allows the comprehender easy elaborations and inferences whenever they are required. It is enough for the well-informed reader to feel that the potential for the elaboration of meaning is there—there is no need to realize it. It is of course possible to do so, and we often do so, sometimes readily, sometimes with the expenditure of considerable effort. Indeed, the deliberate construction of meaning may extend over long time periods, and may be a socially shared activity, for instance, in the case of texts that have special cultural significance, such as the Bible.

This is not a conception of meaning that will make logicians happy. What is true, what is a contradiction, if meaning is subjectively constructed in specific contexts? Logicians had to invent their own sense of meaning. They had to invent logic precisely because the everyday sense of meaning is useless for precise reasoning. Logic, in its various forms, is an extremely successful, well-developed, and useful system. But it is a system that was invented to make precise reasoning possible, not to describe or simulate human cognition. For that we need a very different kind of system, one that is useless for the logician or formal semanticist, but one that meets the needs of the researcher who is interested in describing how humans think, comprehend, decide, and act.

If the meaning of words must be constructed in their context, the difference between literal and metaphorical or idiomatic word meanings is minimized. Both involve constructive processes, and there is no reason to suppose that one kind of construction is necessarily prior to or more difficult than the other. Consider the following examples (Kintsch, 1989):

(a) "*The cat sat on the mat*."
(b) "*He let the cat out of the bag*."

Out of context, there is not much to the understanding of (a): A proposition SIT[CAT,ON-MAT] must be formed, and some more or less dysfunctional associations will be activated, such as *cats purr*, *my cat is black*, or *philosophical argument*. To understand (b) literally, a proposition LET[HE,CAT,OUT-OF-BAG] must be formed, and once again, some random associations having to do with carrying cats around in bags may be activated. To understand (b) as an idiom, the same proposition is formed, but this time it is embedded in a different set of associations, having to do with betraying secrets and surprising revelations of some sort or another. Just how the right set of associates is selected in each case—that is, how the meaning of the phrase is actually constructed in a discourse context—is within the purview of the construction-integration model (Kintsch & Welsch, 1991). Out of context, all three constructions are quite trivial and (except in a linguistic or philosophical discussion) remain superficial. In context, they may be optionally elaborated, depending on the particular context. However, there are no distinct literal or nonliteral processing modes, and it takes people about equally much time to come up with a literal or nonliteral interpretation for such sentences (Glucksberg, Gildea, & Bookian, 1982).

LATENT SEMANTIC ANALYSIS: REPRESENTATION BY VECTORS INSTEAD OF PROPOSITIONS

Propositional representations of text have proven their usefulness in research for over 20 years. Propositional representations of knowledge, as explored earlier, may prove to be equally useful. However, both have a weakness, which is even more serious in practical applications than in research: We cannot construct them automatically and must rely on hand-coding (even though such coding can be reasonably reliable and objective). Hence they are difficult or impossible to use in really large applications. It is fine to analyse propositionally a brief text for an experiment or a simulation, or to construct an illustrative knowledge net to test some empirical implication of a simulation. But one cannot propositionalize a whole textbook, or all of the knowledge of the student studying it. There is, however, another way to represent meaning which is related conceptually to propositional representations and which is not subject to these limitations. It involves a switch in thinking about propositions, not as nodes in a network, but as vectors in a high-dimensional semantic space.

The meaning of a proposition or concept in the abstract is given by its place in a knowledge net. The meaning of a proposition or concept in a discourse context is given by its position in the network representing that discourse, enriched with information retrieved from the knowledge net. Thus, CAT as well as CHASE[DOG, CAT] are defined by the nodes they are linked to in a person's knowledge net. There will be some overlap between these nodes, but also some differences. For example, SCARED[CAT] will be linked only weakly to CAT, but strongly to CHASE[DOG, CAT].

In fact, the labels CAT and CHASE[DOG, CAT] are superfluous; we could equally well denote the two nodes in the network as P_x and P_y. We do not do so, because we would get confused rapidly if we were to use such a denotational system. English-language labels are much easier to remember, but formally they play no role. What is important is the pattern of link strengths to neighbouring nodes in the network.

Furthermore, the graphical notation of a node linked to surrounding nodes in a network is not essential, either. It is equivalent to a vector representation where each row and column corresponds to the nodes of the network and the entry in the ith row and jth column is the strength of the link between nodes i and j.

A concept or proposition can thus be thought of as a vector of numbers, each number indicating the strength with which the concept or proposition is linked to another concept or proposition (Fig. 5).

What determines these numbers? Presumably, they are the end product of life-long experience, of interacting with the world we live in. We learn—by observation, by talking to others, by reading stories—that cats are not normally scared, but that they are when chased by a dog. The values in our vectors, therefore, are the fine-tuned products of numerous and diverse experiences that we have as humans.

There appears to be no way one could provide an artificial system with these numbers by some sort of hand-coding. The system is too complex and too huge and too subtle, and unavailable to reliable introspection. The only way to acquire these numbers is to live a normal human life, learning thorough interaction with the human environment. A machine has no chance to learn all these numbers perfectly, because it cannot live and act like a human—an obvious point, much belaboured by philosophers.

Does that mean that we cannot build a machine that will simulate human cognition adequately? Are we forever restricted to hand-coding of propositions? Not, perhaps, if we shift our criteria somewhat. Machines cannot act and live like humans, and hence they cannot learn from experience as we do, but they can read. So they can learn from reading. A machine that knows about the world only from reading surely is a far cry from a human with real red blood and surging hormones, but there is a lot to be learned from the written word! It is only the second-best choice, but suppose we teach a machine what the

KNOWLEDGE NET:		LATENT SEMANTIC ANALYSIS:	
link to P1	x	dimension 1	x
link to P2	y	dimension 2	y
link to P3	z	dimension 3	z
..	..		..
..	..		..
..	..		..
link to Pn			
..	..	dimension 300	..

FIG. 5. A proposition in a knowledge net is defined by the links that connect it with its neighbours. Analogously, a vector in LSA space is defined by its dimension values.

strength values in all these concept and proposition vectors are by experience with the written word only.

Latent Semantic Analysis (LSA) is a technique that allows us to do something like that (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). LSA uses singular value decomposition, a mathematical generalization of factor analysis. It was originally developed as an information retrieval technique by Deerwester, Dumais, Furnas, Landauer, and Harshman (1990) and extended to discourse analysis and general problems about learning and language by Foltz (1996) and Landauer and Dumais (1997). The reader will have to refer to these sources for an adequate description of the technique and technical details. Here, I only try to give a general impression of what LSA does and then discuss some specific uses of LSA in the investigation of discourse comprehension.

LSA starts its learning process by recording which words occur in the same textual contexts. It reads texts in digital form and counts which words appeared in each segment of a text how many times. Specifically, suppose LSA reads about 4 million words from an encyclopaedia. It counts for each section of the encyclopaedia which words appeared and how often. It makes no inferences about the words it reads, so that "*tree*" and "*trees*" for LSA are different words. For 30,000 encyclopaedia paragraphs containing 35,000 different words, we end up with a huge paragraph by words matrix, with many empty cells since each paragraph contains only a small subset of the words. In the examples we discuss, the results are based on an analysis of this kind provided by Susan Dumais of Bellcore, and described in Landauer and Dumais (1997) and Landauer et al. (1998).

If words were the appropriate units of cognition, we could stop here and define each word as a very long vector, the entries of the vector being the number of times the word has appeared in each paragraph or document. But we know that is not going to be a workable solution: The reason we had to introduce propositional representations in the first place was that words are not satisfactory units for cognition. So instead of defining words directly in terms of documents (and documents in terms of words), LSA substitutes a semantic approximation. It radically reduces the dimension of the space. It does this by the well-known mathematical technique singular value decomposition. A theorem of matrix algebra states that any square matrix M can be decomposed into the product of three matrices:

$$M = A{*}D{*}A'$$

where A and A′ are matrices composed of the eigenvectors of the matrix and D is a diagonal matrix of the eigenvalues (or singular values) of the matrix. In LSA we are interested in non-square matrices, but the theorem generalizes to non-square matrices. The eigenvalues are ordered in terms of their magnitude or importance. Multiplying the three matrices yields back the original matrix, M. What is done in LSA is to throw away most of the eigenvalues (and their associated eigenvectors) and keep only the largest ones, say the 300 largest ones. Multiplying the three matrices thus reduced will not reproduce M precisely, but will only approximate the original M. But that turns out to be a considerable advantage. The original matrix contains too much information—all the details and accidents of word use. By throwing away all this detail, we keep only the essence of each word meaning, its pure semantic structure, abstracted from particular situations. This constructs a semantic space of, in this case, 300 dimensions, in which each word and document from the original matrix can be expressed as a vector. Furthermore, new words and documents can be inserted into this space and compared with each other and with any of the vectors originally computed. There are various ways to compare vectors; the only one I shall discuss here is one closely related to correlation: A measure of relatedness between vectors is the cosine between them in the 300-dimensional space. Identical vectors have a cosine of 1, orthogonal vectors have a cosine of 0, and opposite vectors have a cosine of −1. For instance, "*tree*" and "*trees*" have a cos = .85; "*tree*" and "*cat*" are essentially independent, cos = −.01; "*cat*" and "*The dog chased the cat*" yield a cos = .36.

What is to be gained by this vector representation? Unlike propositional analysis, it is fully automatic and objective. It is computationally not very demanding (once an original semantic space has been constructed). The cosine measure of semantic relatedness is readily interpretable. Thus, it is not necessary to assign links between nodes in a network arbitrarily, or by collecting empirical association data. The cosines between words, sentences, or paragraphs are easily computed and provide objective, empirically based measures. We need no longer guess what the neighbouring nodes of a word (or sentence) are in semantic space—we can look it up in the LSA space. Furthermore, words and documents are treated in exactly the same way. "Documents" for LSA are akin to experiences or episodes for a human learner: The meaning of a word or proposition is determined as much by the episodes in our memory (the documents in LSA) that it is related to as it is by the other words it is related to; that is, meaning involves both semantic and episodic memory. LSA allows us to explore this important aspect of meaning, which is not easily done within conventional approaches.

The initial results that the developers of LSA obtained with this method impressively demonstrate its promise: LSA indeed appears capable of capturing much of word meanings. Landauer and Dumais (1997) found that the vectors for words derived from the encyclopaedia analysis sketched earlier predicted the correct answers to standardized vocabulary tests in which students are asked to judge similarity of meaning. LSA simulations matched the performance of moderately competent students (successful foreign applicants to US colleges). Landauer and Dumais also demonstrated that LSA learned word meanings from reading at about the same rate as late primary school children. Both of these LSA predictions crucially depended on reducing the dimensionality of the semantic space to about 300 dimensions. That is, the words themselves do not matter, but the semantic dimensions derived from their co-occurrences do. In more recent unpublished work by Foltz and by Landauer and Laham, LSA has been trained on introductory psychology textbooks. Then its concept representations were tested with the same multiple-choice tests that students took. LSA usually got about 60% of the items right (the chance level is 22%), somewhere near the 10th percentile of the real students. These results illustrate LSA's impressive ability to approximate human meaning, but also the substantial gap that still exists between humans and LSA. Integrating LSA into the comprehension model developed here might help us to close this gap and at the same time obtain a better and more realistic model.

The vector representation of LSA is similar to the feature vectors popular in many psychological theories, except that we do not have to define, invent, or identify specific features. We need not interpret the values on the 300-dimensional LSA vector (in fact, we cannot), but we can objectively and automatically represent the meaning of various verbal units in this way and use these representations in models of comprehension and memory. The theory of memory based on vector representations is relatively well explored (Estes, 1995) and hence can be used for modelling with LSA vectors. LSA needs no parser; it treats sentences, paragraphs, and whole texts holistically, representing each as a vector. LSA, as currently implemented at least, has its limitations, which need to be explored. The very fact that it needs no parser also means that it does not take into account syntactic information, at least in its present form. LSA allows us to represent global meaning approximately, but not the analytic, formal aspects of human thought. There may be several ways to overcome this limitation, one of which I have begun to explore here: to combine the LSA vector representation with the construction-integration model of comprehension.

LSA is a young technique and the research employing LSA is still in its infancy. Some recent publications suggest its promise, however, such as Landauer and Dumais (1997) and Wolfe et al. (1998). Some further explorations of LSA as a basis for a psychological theory of meaning can be found in Kintsch (1998), the book from which the present paper is an excerpt.

What will be the future of knowledge representations in psychology and artificial intelligence? Hand-coded propositional representations are also useful and have a great deal of intuitive appeal, which makes it very likely that cognitive science will always find some use for propositional representations. When our purpose is to explore or illustrate some point about semantic processes, or to work on small-scale theoretical problems, propositional representations seem ideal, because they are simple and transparent. However, as has been described here, they do not scale up well, and when we need to work on large-scale, practical applications, we require automatic representations of meaning that are both more objective and less cumbersome. LSA may very well prove to be a good choice for such purposes.

REFERENCES

Barclay, J.R., Bransford, J.D., Franks, J.J., McCarrell, N.S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Memory and Language*, *13*, 471–481.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, M.A. Conway, & P.E. Morris (Eds.), *Theories of memory* (pp. 29–101). Hove, UK: Lawrence Erlbaum Associates Ltd.

Collins, A.M., & Quillian, M.R. (1969). Retrieval from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, *41*, 391–407.

Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Estes, W.K. (1995). A general theory of classification and memory applied to discourse processing. In W.A. Weaver, S. Mannes, & C.R. Fletcher (Eds.), *Discourse comprehension* (pp. 35–48). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Foltz, P.W. (1996). Latent Semantic Indexing for text-based research. *Behavior Research Methods, Instruments, and Computers*, *28*, 197–202.

Glucksberg, S., Gildea, P., & Bookian, H.D. (1982). On understanding literal speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, *21*, 85–98.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Kintsch, W. (1989). The representation of knowledge and the use of knowledge in discourse comprehension. In R. Dietrich & C.F. Graumann (Eds.), *Language processing in social context* (pp. 185–209). Amsterdam: North-Holland.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W., & Welsch, D.M. (1991). The construction-integration model: A framework for studying memory for text. In W.E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennett B. Murdock* (pp. 367–385). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Landauer, T.K., Foltz, P., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259–284.

McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 331–343.

Wolfe, M.B., Schreiner, M.E., Rehder, R., Laham, D., Foltz, P.W., Kintsch, W., & Landauer, T.K. (1998). Learning from text: Matching reader and text by Latent Semantic Analysis. *Discourse Processes, 25*, 309–336.