Review TRENDS in Cognitive Sciences Vol.10 No.7 July 2006

Special Issue: Probabilistic models of cognition

Probabilistic inference in human semantic memory

Mark Steyvers[1], Thomas L. Griffiths[2] and Simon Dennis[3]

[1]Department of Cognitive Sciences, University of California, Irvine, California, 92697, USA
[2]Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI, USA
[3]Department of Psychology, University of Adelaide, South Australia, 5005, Australia

The idea of viewing human cognition as a rational solution to computational problems posed by the environment has influenced several recent theories of human memory. The first rational models of memory demonstrated that human memory seems to be remarkably well adapted to environmental statistics but made only minimal assumptions about the form of the environmental information represented in memory. Recently, several probabilistic methods for representing the latent semantic structure of language have been developed, drawing on research in computer science, statistics and computational linguistics. These methods provide a means of extending rational models of memory retrieval to linguistic stimuli, and a way to explore the influence of the statistics of language on human memory.

Rational models of memory

Much of the knowledge that we store in memory is encoded using language. Likewise, most of the stimuli that are used in memory experiments, are linguistic. Language is a dominant element in the environment of human beings, and possesses rich statistical structure. We might thus ask the question of how the statistics of language influence human memory. We will review some recent modeling developments to answer this question from the perspective of rational analysis. Rational analysis is a framework for developing computational models of cognition, making the working assumption that human cognition approximates an optimal response to the computational problems posed by the environment [1,2] (see also Conceptual Foundations Editorial by Chater, Tenenbaum and Yuille in this issue). Rational models emphasize the role of environmental statistics in explanations of human behavior and provide a natural framework in which to explore the interaction between memory and language.

Early rational models viewed the memory system as a predictive system, taking the underlying computational problem to be assessment of the probability that an item needs to be retrieved from memory because of its relevance to the current situation [2–5]. Computation of this 'need probability' is based on two factors: a history factor, which considers the occurrence pattern of the item over time (if the item occurred recently or frequently, it might be needed again soon), and a context factor, which considers the associations between items (if similar items are occurring now, the item might be needed). One of the main contributions of this rational approach has been to show that human memory is remarkably well adapted to the statistics of our environment. For example, Anderson and Schooler showed a close quantitative correspondence between the need probabilities computed from sources such as e-mail subjects and newspaper headings and human performance in list memory experiments. This predictive approach has also turned out to be useful in practical applications such as predicting how users will forage for information on the worldwide web [6].

Much of the original work on rational models of memory emphasized the role of the history factor [2–5]. Several models have recently been proposed that have the potential to extend the rational approach to human memory to better capture the role of context, and in particular the semantic properties of linguistic stimuli. This series of models includes the Retrieving Effectively from Memory (REM) model [7–12], probabilistic topic models [13–16], and the Syntagmatic Paradigmatic (SP) model [17–20]. These models all stress the role of probabilistic inference in memory, and draw on recent techniques from machine learning, statistics, information retrieval, and computational linguistics to provide solutions to the challenges of such inference. For an overview of some of the Bayesian methods relevant to these models, see the Technical Introduction to this special issue by Griffiths and Yuille (Supplementary material online).

The Retrieving Effectively from Memory (REM) model

Early rational models of memory made only minimal assumptions about the way that environmental information is represented in memory and the processing constraints on retrieving this information. The REM memory model (similar to a model by McClelland and Chappell [21]) makes stronger assumptions on the encoding process and representation of information, but continues to emphasize the role of probabilistic inference in explaining human memory. The theory has been

Corresponding author: Steyvers, M. (msteyver@uci.edu).

www.sciencedirect.com 1364-6613/$ - see front matter doi:10.1016/j.tics.2006.05.005

applied to many forms of memory [7,22] but the primary application is recognition memory.

In a recognition memory task, a list of items (e.g. words) is presented to a participant for study. After studying the items, the participant is asked to try to discriminate between old items (words presented on the study list) and new items. The REM model assumes that words are represented by vectors of features, and that each presentation of a word leads to a noisy and incomplete trace of the word vector in memory. Given all the uncertain and incomplete information that is stored in memory, the computational problem is to discriminate between old and new items presented at test. The REM model frames this problem as a calculation of the likelihood ratio that balances the evidence for an 'old' decision against a 'new' decision based on degree of match between the memory probe and contents of memory (see Box 1).

By balancing the evidence for old and new decisions, the model is able to correctly predict 'mirror' effects. Mirror effects occur when an experimental manipulation that improves memory performance is associated with an increase in hit rate (correctly recognizing an old item as old) and a simultaneous decrease in false alarm rates (falsely recognizing a new item as old). For example, a mirror effect for word frequency is typically observed; low frequency words have higher hit rates and lower false alarm rates. The model predicts this effect because more rare features are stored for low frequency words relative to high frequency words. The likelihood calculation takes into account how encoding conditions that improves the diagnosticity of feature matches for low frequency target items also lower the probability of chance matches by low frequency distractor items.

High-dimensional semantic spaces

In the REM model, words are stored as vectors of values representing their phonological, orthographic, and semantic features. In the basic model, these feature values are chosen arbitrarily and do not relate to the actual words presented in a recognition memory experiment. Therefore, the model does not rely on any environmental statistics of word usage, and simply represents an optimal response to the problem of distinguishing new and old words. However, a variety of methods exist that could provide a source of richer representations for words, based on the analysis of large text databases. For example, the Hyperspace Analog to Language (HAL) model represents each word by a vector where each element of the vector corresponds to a weighted co-occurrence value of that

Box 1. Retrieving Effectively from Memory (REM)

REM assumes that items such as words are represented by vectors of discrete feature values. At each presentation of a word, a trace vector of feature values is stored based on a noisy encoding process. At each moment in time, a probabilistic decision is made on whether to store a feature and whether to store the correct feature value (from the complete and noise free vector representing the study word) or a noisy feature value. For each word presented at study, the model stores an incomplete and error-prone trace in memory. A recognition memory decision involves comparison of an item at test to all the stored traces in memory. The data D produced by this comparison consists of all the matches and mismatches D_j of the test item with each trace j. Figure Ia shows an example of a test item being compared with several memory traces.

REM calculates the posterior odds of an item being *old* over *new* by the likelihood ratio of the observed data D times the prior odds for old and new items:

$$\frac{P(old|D)}{P(new|D)} = \frac{P(D|old)}{P(D|new)}\,\frac{P(old)}{P(new)} \qquad \text{(Eqn I)}$$

An *old* decision is made when the posterior odds exceeds some criterion (usually set at 1). The prior odds is set at 1 reflecting the fact that in most recognition memory experiments, the number of old items at test equals the number of new items. Figure Ib shows an example distribution of log posterior odds values for old and new items.

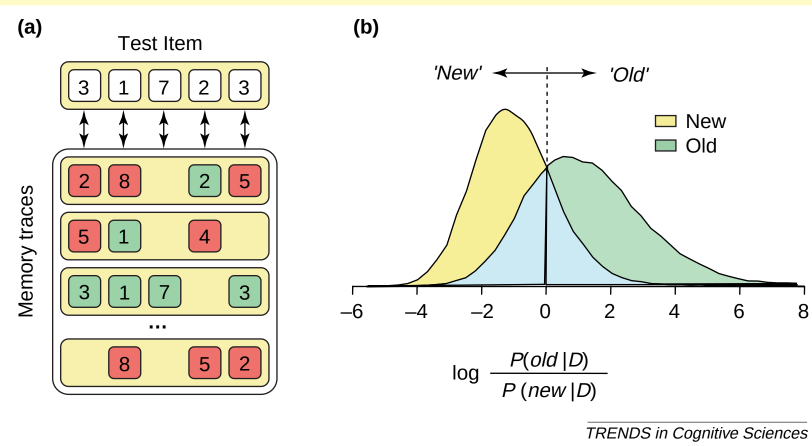

Figure I. **(a)** Example of comparing a test item consisting of five features with a set of noisy and incomplete traces in memory. Green and red boxes show feature matches and mismatches respectively. Missing features reflect positions where no features were stored. The third memory trace provides the most evidence that the test item is old. **(b)** An example distribution of log posterior odds values for old and new items. Note how the old and new distributions cross at log odds of zero – an optimal position assuming there is no bias. Any manipulation in the model that leads to worse memory performance (e.g. decreasing the number of features, or increasing the storage noise or number of traces) will lead to a simultaneous shift in the old and new distribution toward the center preserving the optimal crossover point of the distributions.

word with some other word. The resulting high-dimensional space has been shown to capture neighborhood effects in lexical decision and naming [23].

The Latent Semantic Analysis (LSA) model also derives a high-dimensional semantic space for words but uses the co-occurrence information not between words and words but between words and the passages they occur in [24–26]. The model starts with a matrix of counts of the number of times a word occurs in a set of documents (usually extracted from educational text material), and then applies matrix decomposition techniques to reduce the dimensionality of the original matrix to a much smaller size while preserving as much as possible the covariation structure of words and documents. The dimensionality reduction allows words with similar meaning to have similar vector representations even though they might never have co-occurred in the same document, and can thus result in a more accurate representation of the relationships between words.

Box 2. Probabilistic topic models

A variety of probabilistic topic models have been used to analyze the content of documents and the meaning of words [31,13–15,32]. These models all use the same fundamental idea – that a document is a mixture of topics – but make slightly different statistical assumptions. To introduce notation, we will write $P(z)$ for the distribution over topics z in a particular document and $P(w|z)$ for the probability distribution over words w given topic z. Several topic-word distributions $P(w|z)$ were illustrated in Figure 1, each giving different weight to thematically related words. Each word w_i in a document (where the index refers to the ith word token) is generated by first sampling a topic from the topic distribution, then choosing a word from the topic-word distribution. We write $P(z_i=j)$ as the probability that the jth topic was sampled for the ith word token and $P(w_i|z_i=j)$ as the probability of word w_i under topic j. The model specifies the following distribution over words within a document:

$$P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j) \qquad \text{(Eqn I)}$$

where T is the number of topics. The two terms on the right hand side indicate which words are important for which topic and which topics are important for a particular document, respectively. Several statistical techniques can be used to infer these quantities in a completely unsupervised fashion from a collection of documents. The result is a representation for words (in terms of their probabilities under the different topics) and for documents (in terms of the probabilities of topics appearing in those documents). The set of topics we use in this article were found using Gibbs sampling, a Markov chain Monte Carlo technique (see the online article by Griffiths and Yuille: Supplementary material online; and Refs [15,16]).

The associative semantic structure of words plays an important role in episodic memory. For example, participants performing a free recall task sometimes produce responses that were not on the study list [47]. Typically, these words exhibit strong semantic associations with the words that did appear on the study list [47,48]. To evaluate the effects of semantic association on human memory, word association norms have been developed to measure the associative strength between pairs of words. These norms are typically collected by showing a participant a cue word and asking them to write down the first word that comes to mind. Word association norms exist for over 5000 words, with hundreds of participants providing responses for each cue word [28].

In the topic model, word association can be thought of as a problem of prediction. Given that a cue if presented, what new words might occur next in that context? More formally, the problem is to predict the conditional probability of word w_2 (the response word) given the cue word w_1. The first step in making this prediction is determining which topic w_1 is likely to have been drawn from. This can be done by applying Bayes' rule, with

$$P(z = j|w_1) \propto P(w_1|z = j)P(z = j) \qquad \text{(Eqn II)}$$

It is then possible to predict w_2, summing over all of the topics that could have generated w_1. The resulting conditional probability is

$$P(w_2|w_1) = \sum_{j=1}^{T} P(w_2|z = j)P(z = j|w_1) \qquad \text{(Eqn III)}$$

which we can use to model word association.

Figure I (a) shows the observed and predicted word associations for the word 'PLAY'. Figure I(b) compares the performance of the topic model and LSA in predicting the first associate in the word association norms. The topic model outperforms LSA slightly when either the cosine or the inner product between word vectors is used as a measure of word association.

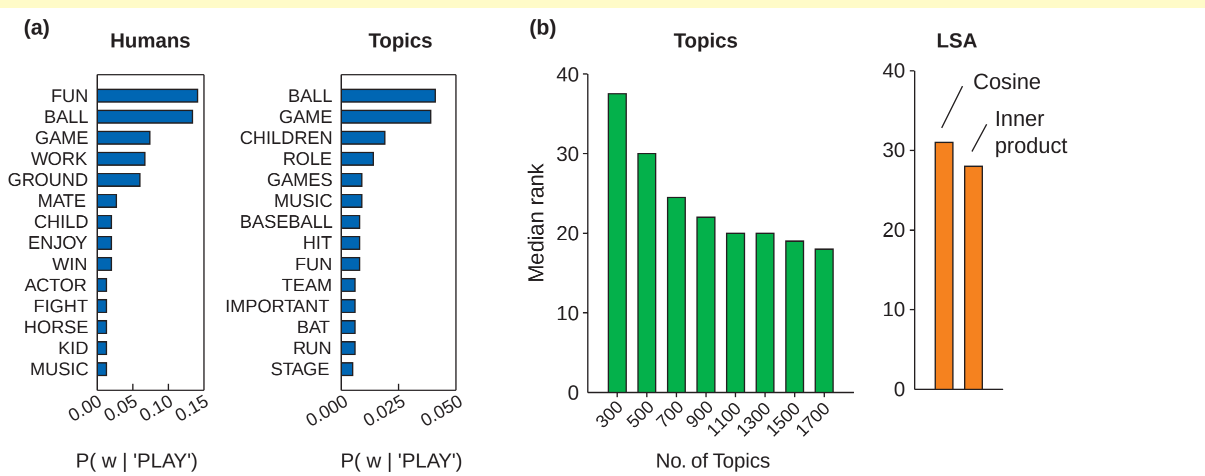

Figure I. **(a)** Observed and predicted response distributions for the word PLAY. The responses from humans reveal that associations can be based on different senses of the cue word (e.g. PLAY–BALL and PLAY–ACTOR). The model predictions are based on a 500 topic solution from the TASA corpus. Note that the model gives similar responses to humans although the ordering is different. One way to score the model is to measure the rank of the first associate (e.g. FUN), which should be as low as possible. **(b)** Median rank of the first associate as predicted by the topic model and LSA. Note that the dimensionality in LSA has been optimized. Adapted from [14,16].

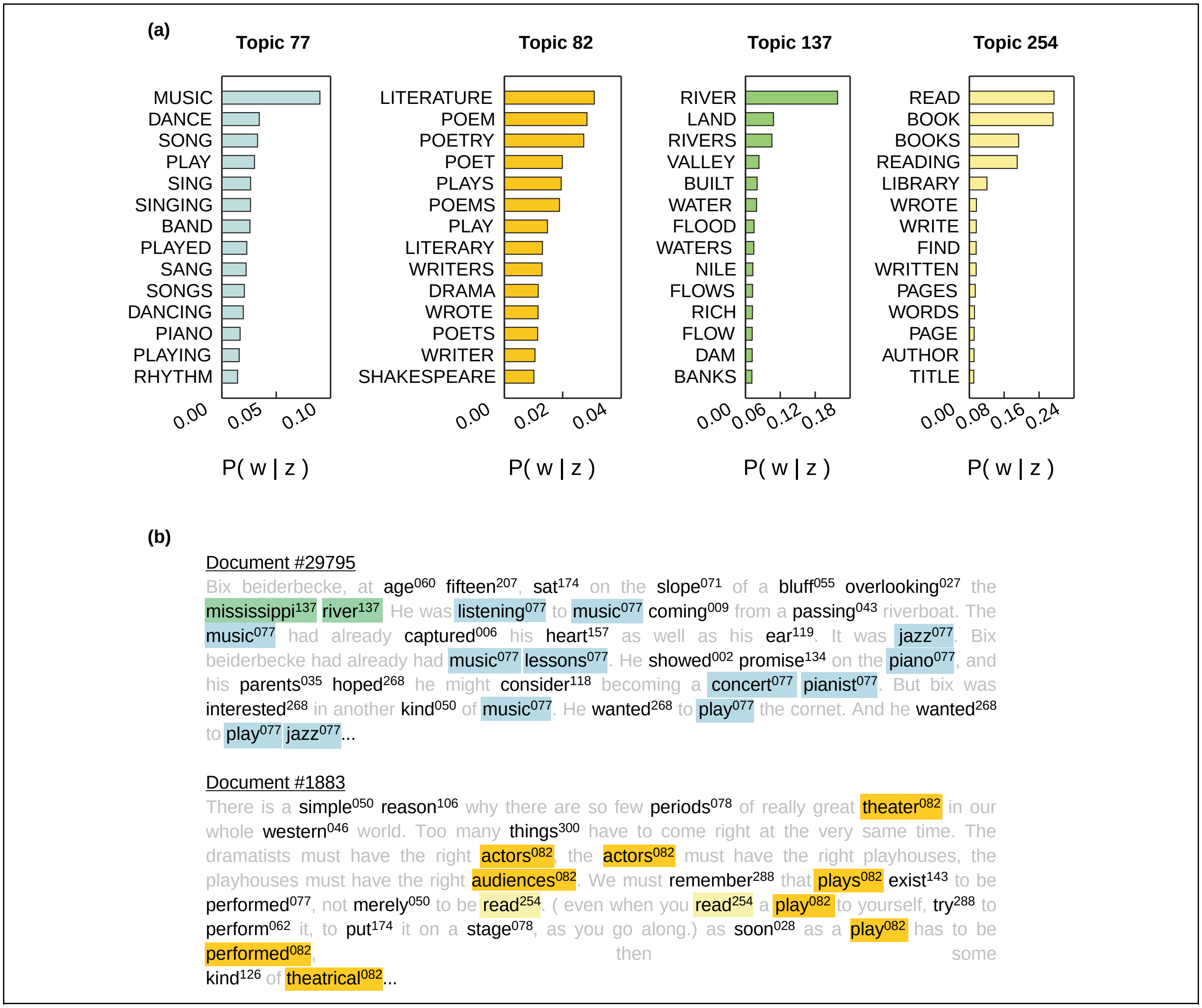

Figure 1. **(a)** Example topics extracted by the LDA model. Each topic is represented as a probability distribution over words. Only the fourteen words that have the highest probability under each topic are shown. The words in these topics relate to music, literature/drama, rivers and reading. Documents with different content can be generated by choosing different distributions over topics. This distribution over topics can be viewed as a summary of the gist of a document. **(b)** Two documents with the assignments of word tokens to topics. Colors and superscript numbers indicate assignments of words to topics. The top document gives high probability to the music and river topics while the bottom document gives high probability to the literature/drama and reading topics. Note that the model assigns each word occurrence to a topic and that these topic assignments are dependent on the document context. For example, the word play in the top and bottom document is assigned to the music and literature/drama topics respectively, corresponding to the different senses in which this word is used. Adapted from [16].

The Word Association Space (WAS) model is another technique for finding representations of words as points in high-dimensional semantic space [27]. Instead of large text databases, the model takes as input a set of word association norms. These norms are formed by asking subjects to produce the first word that comes to mind in response to a given cue [28]. A matrix of associations can be constructed from these data, with the columns being the words used as cues, the rows being the words produced as associates, and the entries in the matrix indicating the frequency with which a word was produced as an associate. The word association space is found by applying the same dimensionality reduction techniques as used in LSA to this matrix. The result is a spatial representation in which words with similar patterns of word associations end up with similar vector representations (even though they might not be directly associated).

The high-dimensional semantic spaces found by LSA and WAS can be used to model semantic effects in episodic memory tasks such as recognition memory and free recall [27,29]. For example, a common finding in free recall is category clustering – words from the same category are often recalled in close temporal succession even though the presentation order of words at study was randomized. Recently, the vector representations found by LSA and WAS have been incorporated into the SAM memory model to explain category clustering and intrusions in free recall [30]. Integrating these representations into rational models such as REM may provide a way to endow the

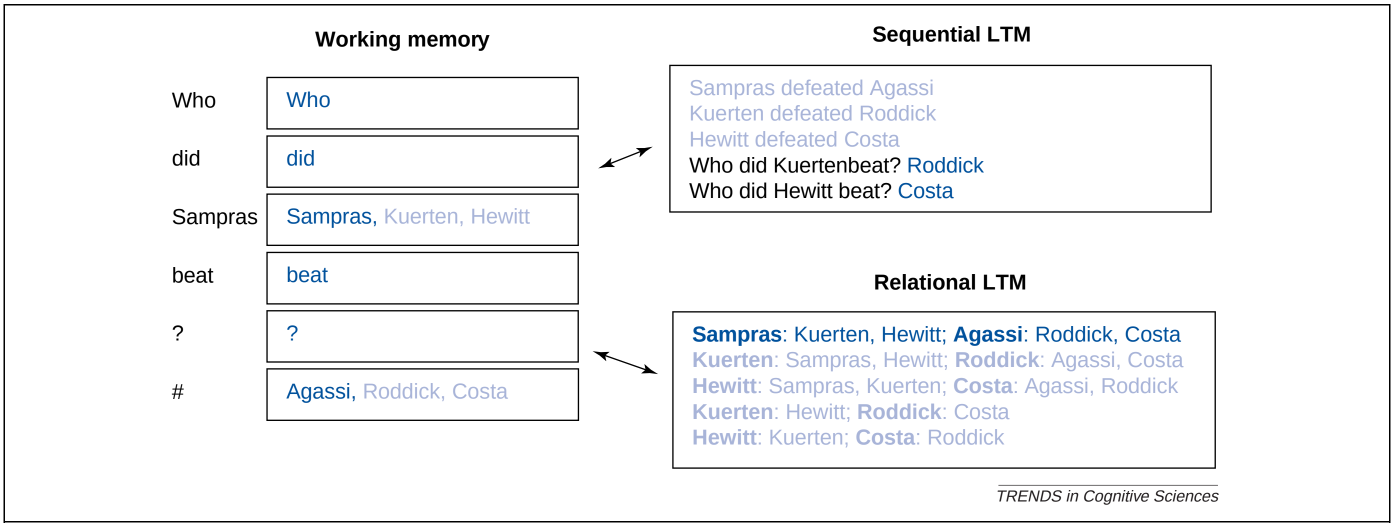

Figure 2. The architecture of the Syntagmatic Paradigmatic (SP) model. Retrieved sequences are aligned with the target sentence to determine words that might be substituted for words in the target sentence. In the example, traces four and five; 'Who did Kuerten beat? Roddick' and 'Who did Hewitt beat? Costa'; are the closest matches to the target sentence 'Who did Sampras beat? #' and are assigned high probabilities. Consequently, the slot adjacent to the '#' symbol will contain the pattern {Costa, Roddick}. This pattern represents the role that the answer to the question must fill (i.e. the answer is the loser). The bindings of input words to their corresponding role vectors (the relational representation of the target sentence) are then used to probe relational long-term memory. In this case, trace one is favored as it contains a binding of Sampras onto the {Kuerten, Hewitt} pattern and the {Roddick, Costa} pattern. Finally, the substitutions proposed by the retrieved relational traces are used to update working memory in proportion to their retrieval probability. In the relational trace for 'Sampras defeated Agassi', 'Agassi' is bound to the {Roddick, Costa} pattern. Consequently, there is a strong probability that 'Agassi' should align with the '#' symbol. The model has now answered the question: it was Agassi who was beaten by Sampras.

rational decision procedure used in the model with a sensitivity to the statistics of language.

Probabilistic topic models

Models such as REM differ from earlier rational models of memory in their construal of the underlying computational problem as one of distinguishing old from new memory traces. Another approach, based on probabilistic topic models, retains the focus on prediction as a central problem of memory, and emphasizes the role of context in guiding predictions. Probabilistic topic models offer an alternative to semantic spaces that is couched in an explicitly probabilistic framework [13–15,31–33]. These models are similar in spirit to LSA; they operate on large databases of text and derive a reduced dimensionality description of words and documents. However, instead of representing words as points in multi-dimensional spaces, probabilistic topic models represent the latent structure of words using topics.

Topic models are based upon the idea that documents are mixtures of topics, where a topic is a probability

Box 3. The Syntagmatic Paradigmatic model

The SP model characterizes sentence processing as a memory retrieval task. As with the REM model described earlier, it assumes that the probability of retrieval of both sequential and relational traces is determined by a likelihood calculation (see Equation II in Box 1). Unlike the REM model, however, the traces in sequential and relational memory are not vectors of features. Rather, they are strings of words and sets of role-filler bindings, respectively. In this box, we describe how strings are compared, but the retrieval of relational traces proceeds in an analogous fashion.

The model uses String Edit Theory (SET) to characterize the similarity of strings of words [49]. As the name suggests, the purpose of string edit theory is to describe how one string, which could be composed of words, letters, amino acids etc., can be edited to form a second string. That is, what components must be inserted, deleted or changed to turn one string into another.

As an example, suppose we are trying to align the sentences 'Sampras defeated Agassi' and 'Kuerten defeated Roddick'. The most obvious alignment is that which maps the two sentences to each other in a one to one fashion:

Sampras	defeated	Agassi
|	|	|
Kuerten	defeated	Roddick

In this alignment, we have three edit operations. There is a **change** of 'Sampras' for 'Kuerten', a **match** of 'defeated' and a **change** of 'Agassi' for 'Roddick'. Using SET, sentences do not have to be of the same length to be aligned. If we add 'Pete' to the first sentence, we can use a **delete** to describe one way in which the resulting sentences could be aligned:

Pete	Sampras	defeated	Agassi
|	|	|	|
—	Kuerten	defeated	Roddick

The '–' symbol is used to fill the slot left by a deletion (or an insertion) and can be thought of as the empty word. While these alignments may be the most obvious ones, there are many other options. But not all of these alignments are equally likely. A mechanism that produces alignments of sentences should favor those that have many matches and should penalize those that require many insertions and deletions. To capture these intuitions, edit operations are assigned probabilities. Typically, match probabilities are higher than change probabilities which are higher than insertion or deletion probabilities. Assuming conditional independence of the edit operations, the probability of an alignment is the multiplication of the probabilities of the edit operations of which it is comprised. Each alignment is an exclusive hypothesis about how the two strings might be aligned and so the probability that the strings are aligned in one of these ways is the addition of the probabilities of the alignments. One set of edit probabilities exist for the case where two strings match and one for the case where they do not. The likelihood ratio is then the ratio of the probability of the correspondence between the target and the trace under the match and don't match models, respectively.

Box 4. Questions for future research

- Can the different computational models for semantic memory be placed in a comprehensive framework of human memory?
- Many of the memory models discussed in this paper are influenced by ideas that were originally developed by computer scientists interested in information retrieval. What lessons can we learn from human cognition that might be relevant to research in information retrieval?
- Can models such as the SP model that extract a propositional representation at the sentence level be extended to capture the structure of propositions at the document level?
- Can topic models, which extract a representation of the gist of a document, be extended to capture semantic structure at the sentence level?
- How could we capture structure at higher levels of abstraction in linguistic stimuli such as narrative structure, rhetorical structure or propositional structure?
- Many computational corpus-based methods make the simplifying assumption that the available text is segmented into thematically coherent contexts. What statistics allow people to divide continuous event streams into contexts?
- Can rational models be extended to capture the interaction between different sources of information, such as episodic and semantic information, or information from different sensory modalities?
- A common finding in memory research is that encoding and retrieval is often guided by the gist as well as the particular details of the to-be-remembered information. Can rational theories of memory be developed that explain the interplay between memory for abstracted information and memory for detail?

distribution over words (see Box 2). The content of a topic is expressed by the probabilities of the words within that topic. A topic model is a *generative model* for documents: it specifies a simple probabilistic procedure by which documents can be generated. To make a new document, one chooses a distribution over topics. Then, for each word in that document, one chooses a topic at random according to this distribution, and draws a word from that topic. Standard statistical techniques (such as the MCMC and EM algorithms discussed in the article by Griffiths and Yuille in this special issue) can be used to invert this process, inferring the set of topics that were responsible for generating a collection of documents. Figure 1a shows four example topics that were automatically derived from the TASA corpus, a set of excerpts from educational materials. Figure 1b shows how the same word but with two different meanings is assigned to different topics based on the document context.

The representations assumed by topic models have several advantages over semantic spaces. In particular, most topics that are learned by the model are individually interpretable, providing a probability distribution over words that picks out a coherent cluster of correlated terms. This is different from semantic spaces, in which individual dimensions are often uninterpretable, and provides a simple way to understand the current state of a learner's beliefs about the semantic context. More generally, topic models provide a means of combining the emphasis on prediction that motivated early rational models of memory with the notion of extracting semantic representations from the statistics of language that has guided approaches such as LSA and WAS. As a generative model for text, topic models can be used to make predictions about which words are likely to appear next in a document or conversation, based on the previous words. The topic structure found by the model is a low-dimensional representation that can be used for making these predictions. Probabilistic topic models thus provide a rational model of how context should influence memory, complementing the focus on the role of history in earlier models, with the effects of context being appropriately modulated by the statistics of the environment. Recent work has focused on extending these models to capture richer semantic structures, such as hierarchies [34], and the interaction between syntax and semantics in statistical models of language [35].

Modeling semantic memory at the sentence level

Many probabilistic models of lexical semantics, including the simple topic models described above, make the simplifying assumption that word order can be ignored. While the success of these models despite this assumption is certainly instructive, it is clear that many aspects of meaning are determined by linguistic structure. In English, in particular, relational information about how roles and fillers combine to create specific factual knowledge is determined to a large degree by the order in which words appear. In keeping with the philosophy outlined above, we define a rational model of human propositional memory; the Syntagmatic Paradigmatic (SP) Model, by specifying a simple probabilistic model of knowledge and allowing representational content to emerge in response to the structure of the environment. The SP model has been used to account for several phenomena including the extraction of lexical information (syntactic, semantic and associative) from corpora, syntactic structure [19], long term grammatical dependencies and systematicity [18], sentence priming [36], verbal categorization and property judgment tasks [18], serial recall [20], and relational extraction and inference [17–19].

The SP model assumes that structural and propositional knowledge can be captured by syntagmatic associations, between words that follow each other (e.g. run – fast), and paradigmatic associations, between words that fit in the same slots across sentences (e.g. deep – shallow). Sets of syntagmatic associations are combined to form structural traces that correspond to individual sentences, while sets of paradigmatic associations are combined to form relational (or propositional) traces that correspond to the same sentences (see Figure 2). The probabilistic model assumes that both structural and relational exemplars from memory are sampled and then stochastically edited to produce observed sentences (see Box 3).

To test the ability of the SP model to capture propositional content, sixty nine articles were taken from the Association of Tennis Professionals (ATP) website and provided to the model [17]. Then 377 questions of the form '*Who won the match between X and Y? – X*' were created. Each question was presented with the final answer slot vacant. On 67% of occasions the model correctly returned the winner of the match. 26% of the time it incorrectly produced the loser of the match. 5% of the time it responded with a player other than either the winner or loser of the match and on 3% of occasions it committed a type error, responding with a word or punctuation symbol that was not a player's name. Furthermore, the model exhibited 'inference by coincidence' in which overlap of role vectors resulted in the model deducing correct answers from indirectly related facts (e.g. Sampras beat Agassi is deduced from the headline '*Sampras wins 12th Grand Slam title*'). The key contribution is that the model achieves this without reference to predefined grammatical, semantic role or inferential knowledge.

Conclusion

Despite their differences in the formulation of the fundamental problem being solved in human memory, the models we have discussed in this paper have two properties in common: the idea that memory should be approached as a problem of probabilistic inference, and a search for richer representations of the structure of linguistic stimuli. We suspect that both the idea and the search will endure. Probabilistic inference is a natural way to address problems of reasoning under uncertainty, and uncertainty is plentiful when retrieving and processing linguistic stimuli. While we view each of the models we have discussed as contributing to the question of how linguistic stimuli might be represented in memory, there is plenty of work still to be done (see also Box 4, and Editorial 'Where next?' in this issue). Contemporary work in linguistics explores richer representations than those assumed in any of these models [37], and accounts of language emphasizing the roles of embodiment and perceptual grounding [38–40] suggest that we will not be able to learn accurate representations of words simply by using text corpora. The availability of resources such as WordNet [41] also provides the opportunity to explore semantic representations that incorporate some of the basic relationships that hold among words.

Humans are remarkably sensitive to the statistical regularities of their environment as demonstrated in areas such as language acquisition [37,42,43], multimodal language learning [38], object perception [44], contour grouping [45] and eye-movements [46] (and the other articles in this special issue give many more examples). Rational analysis provides a natural framework to understand the tight coupling between behavior and environmental statistics. Linguistic stimuli form an excellent testing ground for rational models of memory, since large text corpora can be used to obtain a very good idea of the statistics of the linguistic environment. By developing probabilistic models that can capture these statistics, we can begin to explore some of the factors that influence the structure of human semantic and episodic memory, and perhaps determine the extent to which some of the most basic aspects of cognition are affected by our environment.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tics.2006.05.005

References

1 Anderson, J.R. (1990) *The Adaptive Character of Thought*, Erlbaum
2 Chater, N. and Oaksford, M. (1999) Ten years of the rational analysis of cognition. *Trends Cogn. Sci.* 3, 57–65
3 Anderson, J.R. and Schooler, L.J. (1991) Reflections of the environment in memory. *Psychol. Sci.* 2, 396–408
4 Anderson, J.R. and Schooler, L.J. (2000) The adaptive nature of memory. In *Handbook of Memory* (Tulving, E. and Craik, F.I.M., eds), pp. 557–570, Oxford University Press
5 Schooler, L.J. and Anderson, J.R. (1997) The role of process in the rational analysis of memory. *Cogn. Psychol.* 32, 219–250
6 Pirolli, P. (2005) Rational analyses of information foraging on the web. *Cogn. Sci.* 29, 343–373
7 Shiffrin, R.M. (2003) Modeling memory and perception. *Cogn. Sci.* 27, 341–378
8 Shiffrin, R.M. and Steyvers, M. (1997) A model for recognition memory: REM: Retrieving Effectively from Memory. *Psychon. Bull. Rev.* 4, 145–166
9 Shiffrin, R.M. and Steyvers, M. (1998) The effectiveness of retrieval from memory. In *Rational Models of Cognition* (Oaksford, M. and Chater, N., eds), pp. 73–95, Oxford University Press
10 Malmberg, K.J. *et al*. (2004) Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by Midazolam. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 540–549
11 Malmberg, K.J. and Shiffrin, R.M. (2005) The 'one-shot' hypothesis for context storage. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 322–336
12 Criss, A.H. and Shiffrin, R.M. (2004) Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychol. Rev.* 111, 800–807
13 Griffiths, T.L. and Steyvers, M. (2002) A probabilistic approach to semantic representation. In *Proc. 24th Annu. Conf. Cogn. Sci. Soc.* (Gray, W. and Schunn, C., eds), pp. 381–386, Erlbaum
14 Griffiths, T.L. and Steyvers, M. (2003) Prediction and semantic association. In *Neural Information Processing Systems Vol. 15*, pp. 11–18, MIT Press
15 Griffiths, T.L. and Steyvers, M. (2004) Finding scientific topics. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5228–5235
16 Steyvers, M. and Griffiths, T.L. Probabilistic topic models. In *Latent Semantic Analysis: A Road to Meaning* (Landauer, T. *et al.*, eds), Erlbaum (in press)
17 Dennis, S. (2004) An unsupervised method for the extraction of propositional information from text. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5206–5213
18 Dennis, S. (2005) A Memory-based theory of verbal cognition. *Cogn. Sci.* 29, 145–193
19 Dennis, S. Introducing word order in an LSA framework. In *Latent Semantic Analysis: A Road to Meaning* (Landauer, T. *et al.*, eds), Erlbaum (in press)
20 Dennis, S. and Kintsch, W. The text mapping and inference generation problems in text comprehension: Evaluating a memory-based account. In *Higher Level Language Processes in the Brain: Inference and Comprehension Processes* (Perfetti, C.A. and Schmalhofer, F. eds) (in press)
21 McClelland, J.L. and Chappell, M. (1998) Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. *Psychol. Rev.* 105, 724–760
22 Wagenmakers, E.J.M. *et al*. (2004) A model for evidence accumulation in the lexical decision task. *Cogn. Psychol.* 48, 332–367
23 Lund, K. and Burgess, C. (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208
24 Landauer, T.K. and Dumais, S.T. (1997) A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240
25 Landauer, T.K. *et al*. (1998) Introduction to latent semantic analysis. *Discourse Process.* 25, 259–284
26 Landauer, T.K. (2002) On the computational basis of learning and cognition: Arguments from LSA. In *The Psychology of Learning and Motivation (Vol. 41)* (Ross, N., ed.), pp. 43–84, Academic Press

27 Steyvers, M. *et al*. (2004) Word association spaces for predicting semantic similarity effects in episodic memory. In *Experimental Cognitive Psychology and its Applications* (Healy, A., ed.), pp. 237–249, American Psychological Association
28 Nelson, D.L. *et al.* (1998) *The University of South Florida Word Association, Rhyme, and Word Fragment Norms* (http://www.usf.edu/FreeAssociation/)
29 Howard, M.W. *et al.* Semantic structure and episodic memory. In *Latent Semantic Analysis: A Road to Meaning* (Landauer, T. *et al.*, eds), Erlbaum (in press)
30 Sirotin, Y.B. *et al.* Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychon. Bull. Rev.* (in press)
31 Blei, D.M. *et al*. (2003) Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022
32 Hofmann, T. (1999) Probabilistic latent semantic indexing. In *Proc. 22nd Intl. Conf. Res. Dev. Inf. Retriev.* (SIGIR'99), pp. 50–57, ACM
33 Steyvers, M. *et al.* (2004) Probabilistic author-topic models for information discovery. In W. Kim, R. Kohavi, J. Gehrke, and W. DuMouchel (Eds.) *The Tenth ACM SIGKDD Intl. Conf. Knowledge Discovery Data Mining*, pp. 306–315, ACM
34 Blei, D.M. *et al*. (2003) Hierarchical topic models and the nested Chinese restaurant process. In *Advances in Neural Information Processing Systems 16* (Thrun, S. *et al*., eds), pp. 17–24, MIT Press
35 Griffiths, T.L. *et al*. (2005) Integrating topics and syntax. In *Advances in Neural Information Processing Systems 17* (Saul, L.K. *et al*., eds), pp. 537–544, MIT Press
36 Harrington, M. and Dennis, S. (2003) Structural priming in sentence comprehension. *Proc. 25th Annu. Conf. Cogn. Sci. Soc.* (Alterman, R. and Kirsh, D., eds), pp. 510–515, Erlbaum
37 Pustejovsky, J. (1995) *The Generative Lexicon*, MIT Press
38 Yu, C. *et al*. (2005) The role of embodied intention in early lexical acquisition. *Cogn. Sci.* 29, 961–1005
39 Roy, D. (2005) Grounding words in perception and action: computational insights. *Trends Cogn. Sci.* 9, 389–396
40 Barnard, K. *et al*. (2003) Matching words and pictures. *J. Mach. Learn. Res.* 3, 1107–1135
41 Fellbaum, C. and Miller, G. (1998) *WordNet: An Electronic Lexical Database*, MIT Press
42 Newport, E.L. and Aslin, R.N. (2004) Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48, 127–162
43 Newport, E.L. *et al*. (2004) Learning at a distance: II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* 49, 85–117
44 Fiser, J. and Aslin, R.N. (2005) Encoding multi-element scenes: Statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 134, 521–537
45 Geisler, W.S. *et al*. (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41, 711–724
46 Najemnik, J. and Geisler, W.S. (2005) Optimal eye movement strategies in visual search. *Nature* 434, 387–391
47 Deese, J. (1959) On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22
48 Roediger, H.L. *et al*. (2001) Factors that determine false recall: A multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407
49 Sankoff, D. and Kruskal, J.B. (1983) *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, Addison-Wesley