REM-II: A Model of the Developmental Co-Evolution of Episodic Memory and Semantic Knowledge

Shane T. Mueller
Indiana University
Department of Psychological and Brain Sciences
stmuelle@indiana.edu

Richard M. Shiffrin
Indiana University
Department of Psychological and Brain Sciences
shiffrin@indiana.edu

***Abstract*— Episodic memories are formed from the interpretation of events by semantic knowledge, while semantic knowledge is formed by the accumulation of episodic memories. Through this two-way process, our extensive episodic memory for events in the past co-evolves with our vast knowledge about the world. We present REM-II, a new bayesian account of episodic and semantic memory that explicitly models the development of these two aspects of our long-term memory. REM-II encodes episodic traces as sets of features with different values, and semantic knowledge as a set of co-occurrences of these features. The use of feature co-occurrence allows polysemy and connotation of meaning to be encoded within a single structure, and begins to approach the complexity of human knowledge. We demonstrate knowledge formation in REM-II and show the emergence of semantic spaces through experience and the resultant polysemy and biasing of encoding that REM-II produces.**

Appearing in the *Proceedings of the International Conference on Development and Learning, 5*, 2006.

***Index Terms*— Episodic Memory, semantic memory, memory development**

I. Introduction

A fundamental question in the study of human memory is how new events are interpreted through prior knowledge to form meaningful episodic memories, while at the same time accruing to form this knowledge. A depiction of this interactive process is shown in Figure 1.

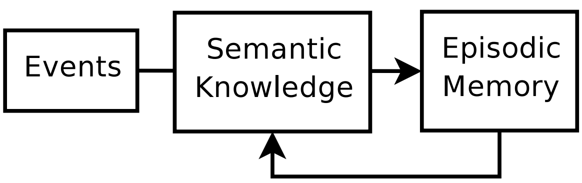

Fig. 1. Events are perceived and interpreted through past knowledge to form episodic memories; these episodes accrue to form this semantic knowledge.

Understanding the developmental evolution of these two memory systems is fundamental for understanding each memory system in isolation. Yet past investigation of both episodic or semantic memory has occurred with little notice of the other, nor of how the two systems are connected or of how they form.

A. *Previous Research on Episodic and Semantic Memory*

Past research has shown there to be intrinsic connections between episodic and semantic memory. For example, [1] and [2] provide classic demonstrations that the basic semantic gist of our experiences is retained, even when the verbatim details are forgotten or inaccessible. Although explicit models exist of episodic memory and semantic memory in isolation, few formal approaches have attempted to integrate the these two memory systems in a comprehensive way.

1) Episodic Memory Research: Numerous models have attempted to understand the storage and retrieval properties of episodic memory (e.g., [3] [4]). Such models have revealed important properties of our episodic memory systems: episodic memories are encoded without great detail, they are durable, and they can be difficult to access without the proper cue. Yet models of episodic memory rarely place this research in the context of the semantic knowledge required to encode episodic memory, either ignored semantic knowledge entirely, or making simplistic assumptions about its nature. Furthermore, models of episodic memory have failed to address how episodes accrue to form knowledge.

2) Semantic Memory Research: Similarly, there have been numerous attempts to understand representations and dynamics of semantic memory [5] [6]. Such models have typically used networks or feature sets to represent the interconnectedness of knowledge, focusing on the organization and structure of knowledge representation [7] [8]. Other research has attempted to construct realistic semantic spaces based on the analysis of text corpora (e.g., LSA [9]; HAL [10]; the Topics Model [11]) in order to develop explicit and meaningful representations of words. Yet few models have attempted connect this work with research on episodic memory, or have examined how the two systems co-evolve with experience.

Next, we will describe work we have done that attempts to connect these two previously distinct lines of research. This work is embodied in the REM-II model, which contains structures related to both episodic and semantic memory, and makes explicit how these systems interact and co-evolve.

II. REM-II: A model of the co-evolution of episodic and semantic memory

In order to investigate the interactions and co-evolution of episodic and semantic memory, we have developed REM-II, a bayesian feature-based model that incorporates properties of REM [4], an earlier model of episodic memory. We augment these properties with detailed assumptions about the nature of semantic knowledge, and the connections between and co-evolution of these two memory systems. We will begin by describing how REM-II represents memories.

A. Memory Representations

1) Episodic Memory Traces: Consistent with the original REM model [4], we assume that episodes are impoverished representations of events in the world, consisting of a sampling of the relevant features with different values, and supported by both bottom-up perceptual features and top-down inferential and relational features. The original REM used as its representation a discrete-valued vector: $[1|2|\;|3|\;|2]$, where each (possibly empty) slot represented a feature (e.g., color, shape, size, etc.), and the numerical value in that slot represented a discrete feature value (e.g., red, blue, round, large, etc.). One limitation of such a representation is that features values can either be present or absent, but there is no way to represent differing levels of importance, and no simple way to add two related vectors into a unitary memory trace, because the two vectors may each have values on some of the same features. Consequently, we have augmented this basic episodic memory representation to allow different features to have different values and strengths in a single trace. This new representation most easily envisioned as a vector where different values of a feature are placed together in the representation, and separated by the $|$ symbol. For example, the vector $[1|2|\;|3|\;|2]$ would be coded as $[1|0\text{-}1|\;|\;|0\text{-}0\text{-}1|\;|0\text{-}1]$, with position within a set of associated features taking the place of the value for a feature, and the value enabling different strengths to be represented, as in: $[3|0\text{-}4|1|1|0\text{-}1\text{-}0\text{-}8|\;|0\text{-}2]$. Such an encoding would represent that the first, second, and fifth feature are relatively more important than others, and that the fifth feature itself has one value that is dominant and a secondary value that is present as well. Such a representation allows feature strength to be encoded, and gives structure to encoding errors, as we assume that errors in encoding are made by encoding the wrong value of a sampled feature from a knowledge matrix.

2) Semantic Knowledge Traces: At first glance, the vector representation we use for episodic memories may appear sufficient to maintain extensive semantic knowledge about a concept. This knowledge could be formed by summing experienced episodes together, filling in missing features and making frequent features stronger than infrequent ones. However, such a representation may be insufficient to capture some of the complexities and breadth of knowledge humans have about concepts in general.

For example, many concepts have multiple connotations and different aspects of their meaning. The word "kitchen" is both a room in a house and (more importantly) a place where cooking and eating happens. Multiple aspects of differing importance can be captured with a feature-vector system, as demonstrated by HAL [10] and the Topics model [11], but we argue that polysemy and connotation go beyond assigning importance to different features: they include the knowledge of which sets of features occur together. For "kitchen", features associated with rooms (like the type of entrance, windows, flooring, etc.) are likely to be encoded together, but not necessarily with its features related to eating, (food, plates, bowls, etc.). Even though the generic concept of a kitchen may contain all of these properties, these two aspects of kitchen share little in common, and a feature-vector representation cannot capture this fact. Unless each connotation of a concept is represented by a single feature, a system using feature-vector knowledge representation will be unable to encode just one sense of the concept, and will instead mix features from multiple contexts together indiscriminately.

In REM-II, we assume that connotation and aspect of meaning are encoded by storing feature co-occurrences of features in the environment. For example, if we examined all encoded episodic traces of "kitchen", we would find that kitchen was sometimes coded in terms of its eating features, while at other times it was coded in terms of its room-like features. To encode feature co-occurrences, we ask, for example, "What features were a part of kitchen when kitchen included the fork feature?". Adding together all kitchens with fork features will produce a single representation for kitchen conditioned on the presence of forks, and similar conditional representations can be formed for all features present in kitchen. Presumably, kitchen conditioned on fork, spoon, plate, or knife features will resemble one another, but differ somewhat from the kitchen conditioned on wall, floor, doorway, or location features. The entire set of conditional representations forms a square symmetric matrix, and each element of the matrix represents the frequency of co-occurrence of any two features in all the previous episodic traces of kitchen.

A simplified co-occurrence matrix representing the knowledge accrued over episodic experiences with a concept is shown in Table 1. Each row depicts a conditional representation for the concept, and values in each cell depict the number of times a feature associated with a specific row co-occurred with the feature associated with a specific column. By studying the matrix, one can see that the first three rows resemble one another and the last two rows resemble one another. This indicates that whenever any of the first three features was encoded, others of those same features were encoded as well.

TABLE I
DEPICTION OF CO-OCCURRENCE MATRIX REPRESENTING SEMATIC KNOWLEDGE OF A CONCEPT.

14	12	13	3	0
12	29	22	5	1
13	22	33	3	1
3	5	3	16	13
0	1	1	13	18

Co-occurrence matrix for five features of a concept. Values in each cell indicate the number of co-occurrences of a specific feature pair.

By using co-occurrences, we are able to capture aspects of meaning that are lost to a simple vector representation in which episodes would simply be summed together. For a summed composite trace, the resulting vector from Table 1 would have been [14|29|33|16|18], which would not distinguish between the two meanings. If the matrix represented "bank", separate representations for the financial institution and the side of a river would exist, while an unconditional composite trace would simply result in wet money! Thus, REM-II uses a feature co-occurrence matrix to represent the semantic knowledge about a concept, in contrast to episodic traces, for which REM-II uses "flat" unconditional traces.

B. Memory Formation

As discussed in the introduction, the episodic and semantic memory systems are intrinsically related in that each is used to form the other. Next, we will describe how these processes are implemented in the REM-II model

1) Formation of new episodic traces: We assume that perceptual features arising from an event in the world enable the memory system to identify an appropriate knowledge matrix. The system must then form an episodic trace by sampling features from the knowledge matrix. Sampling from knowledge will allow the system to encode inferential knowledge based on past experience (i.e., to interpret a pattern of colors as an object like a cup) as well as to augment the representation with associated features (i.e, to know that cups often hold coffee). We assume that the number of features sampled is related to how strongly the episode is encoded, so that more encoding time leads to a stronger and more detailed episodic representations of a concept. To sample a feature from a knowledge matrix, a row of the matrix must first be selected, followed by a feature from that row. The sampled feature will be added to the current episodic trace, and the sampling process will be repeated. Ultimately, a set of feature values representing an episodic memory trace will accumulate through sampling. We assume that encoding is not perfect, so that there is some probability of erroneously encoding features into the episodic trace, by encoding the value of another value that is selected randomly from that feature's base rate in the environment.

2) Attention, bias, and selective encoding of episodic traces: The sampling process described above requires that a row be selected, followed by a feature from that row. We assume that this two-step selection process is fairly flexible to the needs of task, and will be controlled by both attentional factors and the make-up of the knowledge matrix. With no attentional or semantic biases, the sampling of a row may be based solely on the number of past occurrences of each feature, and features may be selected based solely on the number of features in that row. However, depending on the task, other sampling processes may be reasonable. For example, in order to select specific connotations of a concept, rows may be sampled based upon what features have been previously encoded (rather than upon the feature's incidence in past experience). Alternately, a concept may be encoded in terms of another concept by choosing rows from the knowledge matrix by first sampling a feature from the biasing concept. Because tasks differ in the types of information that need to be extracted from the knowledge structure, this too may influence which features get sampled: Free recall may require participants to extract physical features from a matrix identified by its match with semantic features (in order to produce a response), whereas a priming or naming task may involve accessing the matrix based on the match to physical features and selectively sampling associative features to form a semantic episodic code. Thus, we assume that there is some flexibility in the system's sampling strategy, both in the selection of a row from the matrix and a feature from a row. For present purposes we will restrict ourselves to two row-sampling strategies: *selective encoding*, in which the rows are chosen based on the currently-encoded features, and *biased encoding*, in which the rows are sampled based on the features of another biasing concept.

3) Formation of new knowledge traces: Just as episodic traces are formed through the interpretation of events by knowledge, knowledge in formed by the accumulation of episodic traces. We assume this happens iteratively so that each newly encoded episodic trace gets added back into its related semantic knowledge matrix. This is done by calculating a co-occurrence matrix from individual episodic memory traces, which is computed by taking the outer product of a trace with itself. This matrix is simply added back in to the original knowledge matrix, enhancing the knowledge of the concept with the co-occurrences from the most recent episodic trace.

4) Biasing knowledge formation via local semantic context: If a knowledge matrix contains multiple senses of a concept (as shown in Table 1), specific and biased meanings of that concept can be produced as episodic traces. But how do such polysemous knowledge matrices emerge in the first place? Some aspects certainly are introduced by bottom-up perceptual processes: visual properties of a word are typically experienced together, as are auditory properties, but auditory *and* visual properties are experienced together less often. One lesson from numerous research projects involving corpus analysis [9] [10] [11] is that concepts appearing together tend to be related, which is simply a consequence of how the world is structured. We propose that the memory system incorporates the correlated structure of the environment by adding to the representation of a word features from nearby words. We implement this process by generating a trace representing the local semantic context and adding this trace back into the knowledge matrix for a concept. The process by which we generate a local semantic context is depicted in Figure 2.

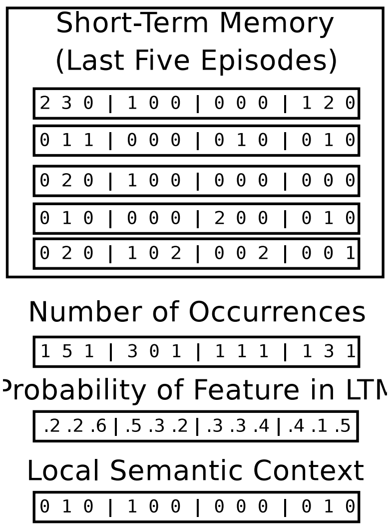

Fig. 2. To generate a local semantic context, we examine recent episodes and find how frequently each feature occurred. We keep only those features that are unlikely to have arisen by chance.

When events are experienced, episodes are created and stored in episodic memory. We assume that the last few episodes experienced form some sort of highly-accessible short-term memory (and for present simulations we use a five-item buffer). To determine which features have been important recently, we examine each feature and determine how many of these traces the feature occurred in (ignoring its strength in any trace), as we want to identify features that have occurred many times recently. Yet this alone is inadequate because some features will have occurred in many of the recent items because they occur frequently in all concepts. So we compare the feature occurrence to the probability of that feature occurring in the environment as a whole, and keep only the relatively unlikely features to form a local semantic context. This process encourages the selection of features that have non-accidental broad support across the local context, encouraging distinct concepts to use distinct feature values.

When an episodic trace is formed, we add this local semantic context to the trace before forming the co-occurrence matrix that gets incorporated into the semantic knowledge matrix. In this way, words that appear in the same context begin to share features representing their relational or associative semantic similarity.

C. *Knowledge Access*

As with the original version of REM, we assume that the memory system accesses stored knowledge by estimating a bayesian likelihood of a target memory trace coming from another probe trace. The present simulations focus on the development of knowledge rather than how it is accessed when memory is probed, and so we will not cover this in great depth here.

Knowledge access is framed as a bayesian probabilistic question: if to determine whether a target trace matches a probe trace, one calculates the probability that the target was encoded from the probe. This is done by inverting the encoding process, which assumed that features were selected from the knowledge matrix and then copied correctly with probability c to the episode. With probability $1 - c$, an error was made, and a feature value sampled from the overall feature base rate was encoded in the episodic trace. To determine whether a target feature coming from a probe feature, we compute a ratio comparing the likelihood that a feature came from the probe versus came because of an error from the base rate. This ratio is:

$$\lambda = \frac{c \cdot probe(cat, feat) + (1 - c) \cdot br(cat, feat)}{br(cat, feat)} \qquad (1)$$

for all features that occur in both the target and the probe. In Equation 1, $probe(cat, feat)$ is the probability of a feature in the probe, and $br(cat, feat)$ is the corresponding probability in the base rate distribution. This ratio is computed for all features categories in which the target and the probe have encoded a value, and the product of these ratios is found to determine the overall likelihood ratio of a target-probe pair. If this ratio is high enough (typically greater than 1.0), the system will determine that the probe and target match.

III. Demonstrations: Development of Knowledge

REM-II makes explicit assumptions about how the semantic and episodic memory systems co-evolve with experience. We introduced a process by which features from a local semantic context get added to semantic representations in order to introduce relational features that code for co-occurrence. In the following demonstrations, we will show several aspects of how this process evolves and how similarity relationships between words can develop from it.

In the following simulation, we use a small feature set (6 feature categories with 3 values per category) and lexicon (eight words) to demonstrate the operation of the model. Each word was given a random initial configuration (shown in the upper left panel of Figure 3), where each column depicts a feature and each row depicts the representation for a specific word. The lexicon was arranged so that there were two primary meaning groups (Group A: Words 1, 2, and 3; and Group B: Words 5, 6, and 7), and two polysemous words that appeared in both Group A and Group B, but never together. We expect that with experience, words from Group A will grow similar to one another, as will words from Group B. Additionally, the polysemous words (4 and 8) will develop representations that are similar to both sub-groups, and similar to one another (even though they never appeared in the same context). Finally, one should be able to bias new episodic encodings of the polysemous words to resemble words from Group A or Group B. In our simulation, each iteration of learning consisted of the following sequence: (1) ten tokens randomly sampled from Group A; (2) ten tokens randomly sampled from Group B; (3) 20 tokens sampled from Group A and polysemous Word 4; (4) 20 tokens sampled from Group B and polysemous Word 4; (5) 20 tokens sampled from Group A and polysemous Word 8; and (6) 20 tokens sampled from Group B and polysemous Word 8. This entire process was repeated 5000 times, and the results of a single typical run are shown.

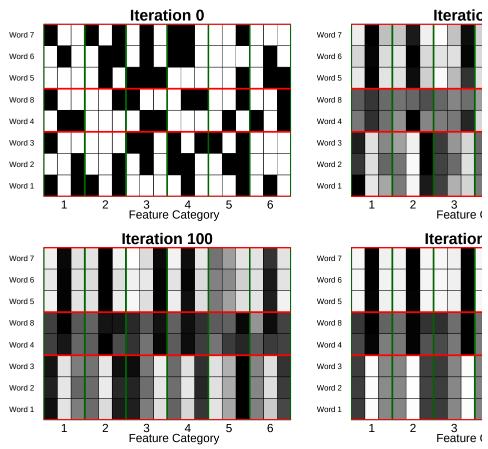

Fig. 3. With experience, the initially random feature representations converge to form similar representations within groups that are distinct between groups. Each row represents the composite matrix representation for a specific word token, while each column is a feature value. Darker cell indicate more occurrences of that feature in the representation.

Results from this simulation are shown in Figures 3 through 4. Figure 3 shows how the representations of words within a single contextual group grow together, while the polysemous words exhibit properties of both groups. Interestingly, the semantic similarity begins to emerge after just a few iterations, and after many thousands of iterations (and thus tens of thousands of experiences with each word), the meanings of different words within each group become indistinguishable.

To visualize the semantic space produced by this process, we computed pairwise dissimilarity matrix for the knowledge traces representing each word.[1] This dissimilarity matrix was then analyzed with the statistics software R using the multi-dimensional scaling [12] function `isoMDS` from the `MASS` library, and the `diana` hierarchical clustering [13] function from the `cluster` library.

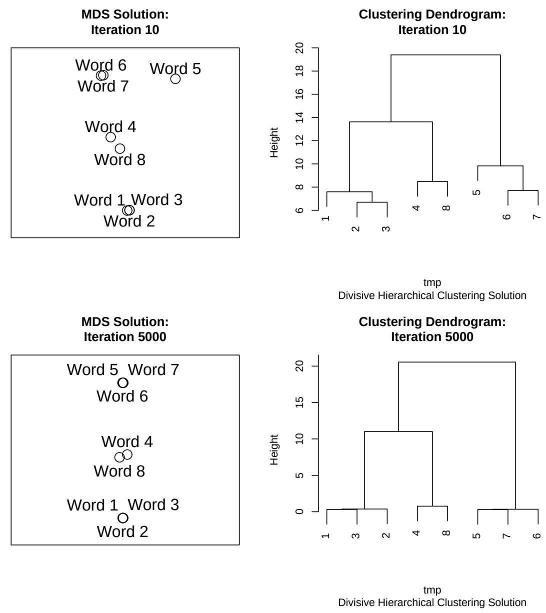

Fig. 4. Similarity space for words early and late in development. Top panel shows space after ten iterations; bottom panel after 5000 iterations. Left panel: Multidimensional-scaling solution; right panel: divisive hierarchical clustering solution. Both solutions show that representations attain within-group similarity and between-group dissimilarity. The polysemous words (4 and 8) fall between these groups, and grow more similar with experience. Open circles represent position in MDS space for words.

The MDS and clustering solutions show that even after relatively few presentations (ten iterations), basic group clustering has begun to emerge, with Group A words (1, 2, and 3) and Group B words (5, 6, and 7) growing close to one another but away from each other, while the polysemous words lie between these two groups, and close to each other.

[1]To compute dissimilarity from a matrix of feature co-occurrences, we divided each row of the matrix by the total number of counts in that row, so that each entry became a probability. We computed a sum squared difference between the probabilities in two matrices to get a total dissimilarity score between matrices. This was done for all pairs of words to produce a complete dissimilarity matrix for the word representations.

As development proceeds, clustering becomes tighter and similar representations become indistinguishable.

This example illustrates that the semantic representations of polysemous words are different from either of the contexts they appear in, and words that never appear together can grow similar by appearing in similar contexts. Yet it remains to be seen whether the knowledge representations for the polysemous words and episodic traces they produce do in fact segregate the two senses of the word. To examine this, we generated episodic traces from the knowledge structures produced after 10,000 iterations of learning.

For this example, we generated sets of 100 episodic traces by sampling from specific knowledge matrices repeatedly (a total of 40 samples per trace). First, traces were generated from Word 1 and Word 5 (prototypical members of Groups A and B, respectively) and from Word 4 and Word 8 (the two polysemous words). For each of these episodes, selective sampling was used, so that new rows were chosen based on previously-encoded features. We also encoded episodes from polysemous concepts in a biased manner: biased either toward a word from Group A or from Group B. Similarity was computed between all 800 encoded episodes by converting the sampled traces to probabilities and finding the sum squared difference between pairs. One global MDS analysis was performed on the resulting dissimilarity matrix, and the results for subsets of the stimuli are shown in Figure 5.

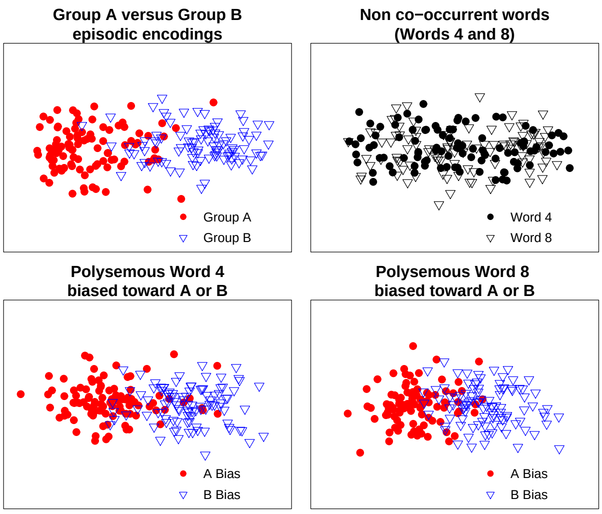

Fig. 5. Select subsets of MDS solution for episodic traces encoded from various knowledge matrices.

The top left panel if Figure 5 shows the position in a two-dimensional similarity space for 100 episodic encodings of a word from Group A and 100 from Group B. There is little overlap between the two distinct groups of points, indicating that episodes encoded from the two matrices attained distinct representations. The top right panel of Figure 4 shows 100 episodic encodings each of the two polysemous words (4 and 8). Although these words never appeared together in the same context, each appeared alone with words from both Group A and Group B. Consequently, over time, they developed similar representations, and episodic traces encoded from each are indistinguishable and cover the entire space of pure encodings from Group A and Group B. Yet when these polysemous words are encoded in a biased fashion, each distinct meaning can be selected from the composite trace. The bottom panel shows how episodes from each polysemous co-occurrence matrix can be produced that are similar to either Group A or Group B, simply by selecting rows from the knowledge matrix based on a word from either of these groups. In fact, there is almost no overlap between the different connotations of the polysemous words, indicating that the aspects are quite distinct.

IV. Conclusion

Semantic and episodic memory have long been considered separate memory systems, and have been studied independently. Yet each is required to form the other, and the process by which the two systems develop is poorly understood. We have presented REM-II, a new model that incorporates both episodic memory and semantic knowledge which explicitly addresses how each system is used to form structures in the other system, and how the two systems develop with time to produce our complex and detailed knowledge of the world. REM-II assumes that although episodic traces are relatively impoverished representations of events, the semantic knowledge they accrue to form keeps track of the co-occurrence of features within each knowledge trace. By accumulating feature co-occurrences, connotation of meaning and polysemy represented. Furthermore, we assume that co-occurrence of concepts with other concepts is incorporated into the representation by adding features from a local semantic context into the knowledge trace for a concept. This allows concepts that are similar to develop similar representations, and concepts that appear in distinct contexts to develop distinct representations for these different connotations.

Acknowledgment

This research supported by NIMH grant #12717 to Richard M. Shiffrin. We would like to thank the members of the Memory, Attention, Perception Lab at Indiana University.

References

[1] F. C. Bartlett, Remembering. Cambridge: Cambridge. Univ. Press, 1932.
[2] J. D. Bransford and J. J. Franks, "The Abstraction of Linguistic Ideas," Cognitive Psychology, vol. 2, pp. 331-350, 1971.
[3] G. Gillund, G., and R. M. Shiffrin, "A retrieval model for both recognition and recall," Psychological Review, vol. 91, pp. 1-67, 1984.

[4] R. M. Shiffrin and M. Stevyers, “A Model for Recognition Memory: REM—Retrieving Effectively from Memory,” Psychonomic Bulletin & Review, vol. 4, pp. 145-166, 1997.

[5] A. M. Collins and M. R. Quillian. “Retrieval time from semantic memory,” Journal of Verbal Learning and Verbal Behavior, 8, pp. 240–247, 1969.

[6] D.E. Meyer and R.W. Schvaneveldt. “Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations,” Journal of Experimental Psychology, vol. 90, pp. 227-235, 1971.

[7] T. K. Landauer and S. T. Dumais. “A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge,”, Psychological Review, vol. 104, 211-240, 1997.

[8] C. Burgess. “From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model,” Behavior Research Methods, Instruments, & Computers, vol. 30, pp. 188-198, 1998.

[9] T. Griffiths and M. Steyvers, “Finding Scientific Topics,” Proceedings of the National Academy of Sciences, 101 (suppl. 1), 5228-5235, 2004.

[10] T. F. Cox and M. A. A. Cox, Multidimensional Scaling, Chapman & Hall, 2001.

[11] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.