BRAIN AND COGNITION 30, 257–434 (1996)
ARTICLE NO. 0018

TENNET VI: Theoretical and Experimental Neuropsychology

May 10–14, 1995

Université du Québec à Montréal, Pavillon Judith-Jasmin, Salle Marie-Gérin-Lajoie
Montréal, Québec, Canada

SYMPOSIUM II

Developing Cognitive and Neural Models of High-Dimensional Semantic Space
(*Organized by Curt Burgess*)

5. *Hyperspace Analogue to Language (HAL): A General Model of Semantic Representation.* KEVIN LUND AND CURT BURGESS, University of California, Riverside.

Semantic information plays a pivotal role during memory retrieval and language comprehension. We present a model of semantic memory that we call the Hyperspace Analogue to Language (HAL) that utilizes a high-dimensional semantic space constructed from a lexical cooccurrence matrix. This matrix was formed by analyzing a 160 million word corpus. Word vectors were then limited to only the most informational vector components. Using these vector representations of words, we present a variety of experiments in which this computationally derived semantic information supports the empirical results of corresponding experiments with human subjects. These experiments include family resemblances and typicality, semantic, episodic, and associative priming, semantic constraints in parsing, vector substitution in TODAM and CHARM memory models, Toglia and Battig's Semantic Word Norms, and more. Our methodology exploits that regularities of language in a large corpus such that the extraction of semantic information is possible. The methodology employed in HAL does not require supervised learning or other system feedback and works on very noisy, speech-like input. A limitation to previous models of semantic processing is that either the semantic representations required extremely time-consuming human judgments about the items or the semantic representations were simply conjectural. HAL is a model that provides for the generation of semantic representations of real words used in real language.

6. *Modeling the Cerebral Asymmetries of Semantic Memory in High-dimensional Space: HAL's Brain.* CURT BURGESS AND KEVIN LUND, University of California, Riverside.

Our computer model of cerebral asymmetries of semantic phenomena utilizes high-dimensional semantic space constructed from a lexical cooccurrence matrix initially employed in our general semantic memory model called the Hyperspace Analogue to Language (HAL). Our high-dimensional semantic space uses 200 dimensions, which were extracted with HAL's cooccurrence matrix. This matrix was automatically constructed from a 160 million word corpus. In order to simulate the cerebral asymmetries found with lexical/semantic stimuli (e.g., most notably the work of Chiarello, Burgess, Beeman, Zaidel, Brownell, and colleagues) we have evolved HAL into a processing, as well as a representational, model by making two crucial processing modifications. Left hemisphere (LH) and right hemisphere (RH) processing are characterized by using an "activation volume." The LH has a smaller activation volume than the RH, resulting in the RH activating more peripheral semantic information than does the LH. This activation volume increases asymptotically over time at about 300 msec in the LH and at about 750 msec in the RH. Furthermore, this LH volume decreases to its resting state by 750 msec, whereas the RH volume slowly decays away. This accounts for how the LH will rapidly select information after the multiple activation of semantic information and how the RH maintains activation for a wider range of information. This approach to modeling hemispheric asymmetries brings together methodologies from computational linguistics and psycholinguistics and has the advantage that the model can learn its "representations" in high-dimensional space from a large corpus of text in an unsupervised fashion.

7. *HAL's Semantics: Lost in Vector Space?* CATHERINE L. HARRIS, BOSTON UNIVERSITY.

The idea of representing words' semantic similarities in terms of closeness in conceptual space dates back to Ross Quillian and William James. The pattern recognition and connectionist work of the 1980s have further increased the familiarity of vector spaces as representations. The innovative aspect of Burgess and Lund's HAL is that vector representations are constructed automatically, with the bulk of the English language included. HAL thus stands as an alternative tool to word association norms and raters' semantic judgments for researchers who want to select or match stimuli. But is HAL a contender as a model of semantic representation? I discuss two types of objections. (1) Like a dictionary, HAL's word meanings are defined entirely by other words. Important questions in semantics are how word meanings are related to concepts, and how novel word combinations are interpreted. It is unclear if HAL will help address these questions. (2) Burgess and Lund made a number of design decisions in order to do computations on a 300 million word corpus: (a) statistics were done with words as units, and (b) the dimensions of semantic space correspond to words. While words are salient units (especially for literate adults), there is considerable evidence that humans do computations on and learn associations between units both smaller and larger than words. More plausible than words as dimensions are conceptual features as dimensions. Connectionist techniques, with a small training corpus, can extract units whose size varies depending on distributional regularities; both semantic and associative regularities are extracted and represented across the same set of processing units. Unfortunately, the problem of defining a training task, and training on 30,000+ words, makes a connectionist alternative to HAL infeasible.