Grossberg and Colleagues Solved the Hyperonym Problem over a Decade Ago.

Jeffrey Bowers

Department of Psychology

Rice University

Houston, Tx 77251-1892

(713) 527-8101

bowers@rice.edu

running head: the hyperonym problem

Abstract

The authors describe a model of speech production in which lemma access is achieved via input from non-decompositional conceptual representations. Part of their motivation for adopting this characterization of concepts is their claim that existing decompositional theories are unable to account for lexical retrieval, due to the so-called hyperonym problem. However, existing decompositional models have solved a formally equivalent problem.

An important theoretical claim of the Levelt et al. is that conceptual knowledge is coded in an non-decompositional format; that is, concepts for the various morphemes in a language are primitive (section 3.1.1). One of the reasons the authors have adopted this position is the claim that existing decompositional accounts have been unable to solve the so-called hypernym problem, which Levelt (1992, p. 6) defined as follows:

> When lemma A's meaning entails lemma B's meaning, B is a hyperonym of A. If A's conceptual conditions are met, then B's are necessarily also satisfied. Hence, if A is the correct lemma, B will (also) be retrieved.

For example, when a speaker wants to express the concept CAT, all the conceptual conditions for retrieving the lemma for ANIMAL are satisfied as well, since the meaning of cat entails the meaning of animal. Thus, both lemmas would be retrieved, contrary to what in fact occurs

In this commentary, I simply want to point out that one well developed network model has in fact solved a formally equivalent problem maintaining a decompositional approach; namely, the adaptive resonance theory (ART) of Grossberg and colleagues (e.g., Carpenter & Grossberg, 1987; Grossberg, 1976).

The Carpenter and Grossberg (1987) Solution

Carpenter and Grossberg (1987) addressed the question as to how a network can correctly categorize subset and superset visual patterns. For illustrative purposes, consider the case of visual word recognition in which the words my and myself are presented. The problem is to develop a model that can correctly categorize both patterns, because as the authors note, the superset pattern myself contains all the features of the subset my, so that the presentation of myself might be expected to lead to the full activation of the lexical orthographic codes of my as well as myself. Similarly, if the features of my are sufficient to access the subset pattern, it might be expected that the superset pattern myself would access the subset pattern as well. Thus, how does the network decide between the two inputs? The hyperonym and problem revisited.

Their solution was embedded within an ART network that contains two fields of nodes, an input and output layer called F1 and F2. The nodes in F1 each refer to a feature (a letter in the above example), so that the nodes that become active in response to the input my are a subset of the active nodes in response to myself. The active nodes generate excitatory signals along pathways to the target nodes in F2, which are modified by the long term memory traces (LTMs) that connect F1 and F2. Each target node in F2 sums up all of the incoming signals, and transforms this pattern of activation based on the interactions among the nodes of F2, resulting in a single active node at F2. See Figure 1.

For present purposes, consider the case in which v1 and v2 refer to the two nodes in F2 that code for the subset and superset patterns in F1, respectively. Thus, the subset pattern at F1 must selectively activate v1, and the superset pattern at F1 must selectivity activate v2. In order to achieve this result, Carpenter and Grossberg (1987) incorporated learning rules that followed the Weber and the Associative Decay principles. Very briefly, according to the Weber rule, there is an inverse relationship between LTM strength and input pattern scale, so that the LTM traces connecting the subset pattern at F1 to v1 are stronger than the LTM traces connecting this same subset pattern to v2 (otherwise, the superset could activate v1). And according to the Associative Decay rule, LTM weights decay towards 0 during learning when nodes at F1 and F2 are not co-active. In particular, LTM weights decay to 0 between inactive nodes in F1 that are part of the superset pattern to v1. Together, these rules accomplish the goal: Since the superset pattern includes more active nodes than the subset, the Weber Law Rule insures that the LTM traces in the pathways to v2 do not grow as large as the LTM traces to v1. On the other hand, after learning occurs, more positive LTM traces project to v2 from F1, which combine to produce larger activation to v2 than to v1 when the superset is presented. Thus, there is a trade-off between the individual sizes of the LTM traces and the number of traces, which allows direct access to both subset and superset representations.

<u>Carpenter and Grossberg's (1987) Proof that the Weber and Associative Decay Rules Achieve Subset and Superset Access</u>

As learning of an input pattern takes place, the bottom-up LTM traces joining F1 and F2 approach an asymptote of the form:

/(ß + |I|) (1)

where and ß are positive constants, and |I| equals the number of nodes active in a pattern I in F1. From (1), larger |I| values imply smaller positive LTM traces in the pathways encoding I.

Direct access to the subset and superset patterns can now be understood as follows.

By (1), the positive LTM traces which code the subset pattern have the size

/(ß + |subset|) (2)

and the positive LTM traces that code the superset have the size

/(ß + |superset|) (3)

When the <u>subset</u> is presented at F1, |subset| nodes in F1 are active. The total input to v1 (T_{11}) is proportional to

$$T_{11} = |subset| / (ß + |subset|) \quad (4)$$

And the total input to v2 (T_{12}) is proportional to

$$T_{12} = |subset| / (ß + |superset|) \quad (5)$$

Because (1) defines a decreasing function of |I| and because |subset| < |superset|, it follows that $T_{11} > T_{12}$. Thus, the subset pattern activates v1 instead of v2.

When the <u>superset</u> is presented to F1, |superset| nodes are active. Thus the total input to v2 (T22) is proportional to

$$T_{22} = |superset| / (ß + |superset|) \quad (6)$$

Now, the Associative Decay Rule is critical, because only those F1 nodes in the superset that are also activated by the subset project LTM traces to v1. Thus, the total input to v1 is proportional to

$$T_{21} = |subset| / (ß + |subset|). \quad (7)$$

Both T22 and T21 are expressed in terms of the Weber function

$$W(|I|) = |I| / (ß + |I|), \quad (8)$$

which is an <u>increasing</u> function of |I|. Since |subset| is smaller than |superset|, $T_{22} > T_{21}$. Thus, the superset activates v2 rather than v1. In summary, direct access to subsets and supersets can be traced to the opposite monotonic behavior of the functions (1) and (8).

In conclusion, the hyperonym problem can be solved within a decompositional framework . Whether nor not a general theory of lexical retrieval can be accomplished using this framework remains to be seen however.

References

Carpenter, G. & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, 37, 54-115.

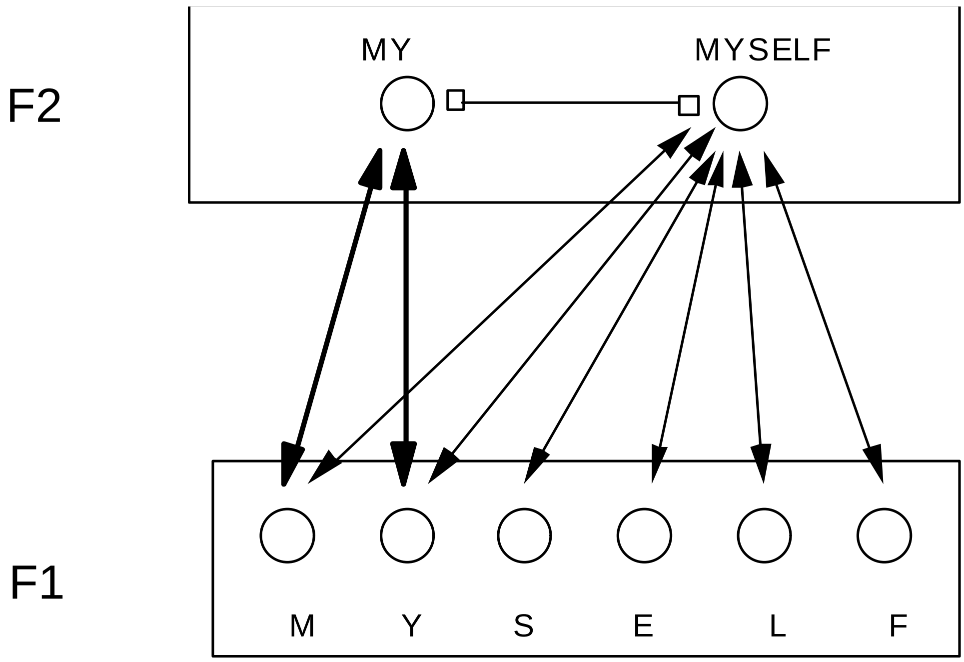
Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics, 23, 187-203

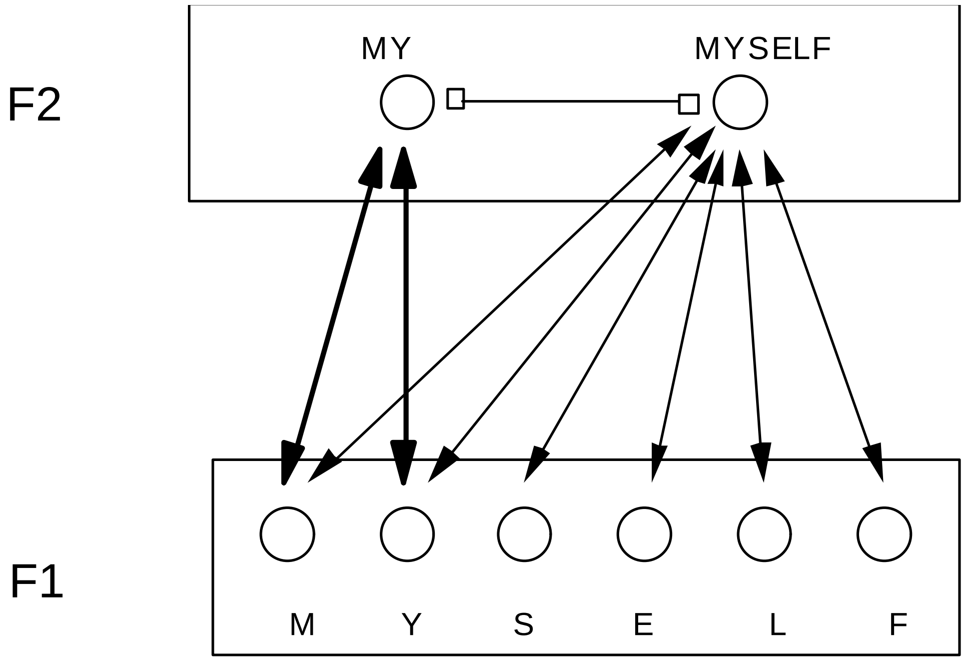
Levelt, W.J.M. (1992). Accessing words in speech production: Stages, processes and representations. Cognition, 42, 1-22.

Figure Captions

1. Direct connections between levels F1 and F2 of ART network. The solid arrows indicate excitatory connections, and the squares indicates inhibitory connections.

Figure 1




F2
MY
MYSELF
F1
M
Y
S
E
L
F