The Human Semantic Potential:
Spatial Language and Constrained Connectionism

Terry Regier
(University of Chicago)

Cambridge, MA: The MIT Press
(Neural Network Modeling and
Connectionism Series, edited by Jeffrey
L. Elman), 1996, xv+220 pp;
hardbound, ISBN 0-262-18173-8, $37.50

Reviewed by
Whitney Tabor
Massachusetts Institute of Technology

1. Introduction

One of the implications of connectionist work on language is that the "language mechanism" is not essentially different from mechanisms that perform other cognitive tasks. This leads naturally to the notion that language learning must be understood in the context of learning about the world more generally. It is difficult to think of how to constrain a modeling exercise that adopts this perspective, for such a system requires an encoding not only of some interesting part of language but also of the corresponding part of "the world." If one were to undertake this rather daunting task, a natural place to start would be with a subdomain of language in which the important semantic contrasts can be expressed in a two-dimensional visual world and to use low-resolution computer graphics for the encoding of this world.

In *The Human Semantic Potential*, Regier takes precisely this approach, by focusing on the domain of closed-class spatial markers (e.g., prepositions such as *above, below, left, right, in,* and *on* in English). Bit map "movies" of a **trajector** that moves in relation to a **landmark** for several movie frames, or that simply exists, are used as the input to a connectionist learning model; the outputs are labels corresponding to prepositions.

The scientific game plan is to construct a model that learns the semantics of spatial markers on the basis of exposure to examples in a given language; the learnable sets of markers under the model then constitute a typology of possible sets of spatial markers in languages of the world. Thus the model makes cross-linguistic typological predictions. Moreover, since the model generalizes from a subset of examples in each language to an assessment of the aptness of each word in all representable conditions, it makes language-internal predictions about the meanings that particular spatial markers can have.

Given this framework, it seems reasonable to assess the project in terms of how much we learn about these two domains of prediction from it. In the former case— cross-linguistic typological prediction—the results are only mildly interesting: only a very small subset of the range of predictions that the model makes is revealed, apparently because of the assumed unanalyzability of the connectionist network at the core of the model. In the case of the second task—generalization from positive examples— the results are more appealing: Regier shows how a linguistically-motivated variation on a standard connectionist learning algorithm provides an effective mechanism for

generalizing in the absence of negative evidence. I'll elaborate on these two assessments after providing a summary of the contents.

2. Summary of the Book

The book opens with a presentation of the modeling project. A discussion, from a cognitive-linguistics perspective, of the contrasting semantics of spatial prepositions in several languages provides motivation for framing the model in terms of certain primitive notions such as **orientation** and **contact**. A readable introduction to connectionist models and the back-propagation learning algorithm is then provided. "Constrained connectionism" is introduced: models consist of some nodes that are specifically designed to pick up on features relevant to the task at hand and some nodes whose role in the representation is determined by the learning process (like the "hidden nodes" of stereotypical connectionist networks). A simulation is described in which the model, on the basis of limited examples, correctly generalizes the meanings of the English prepositions *above, below, left, right, in, out, on,* and *off* without the benefit of explicit negative evidence. Then, the structures built into the network are reviewed in detail: there are **orientation combination** nodes, which measure angles between important reference lines in the current scene (e.g., there is a node that detects the angle between a vertical line and the line connecting the centers of mass of the trajector and landmark) and there are **map comparison** nodes, which detect features such as contact and inclusion. The temporal structure of the scenes is detected by means of nodes that record the maximum, minimum, and average values of other nodes in the network (over the course of the current movie). In this regard, the model diverges from the bulk of connectionist models of temporal processing, which seek to solve in general the problem of detecting correlations between events that are spread out over time, rather than positing task-specific detectors. Helpfully, Regier provides a comparison with some of these other approaches. A sample run-through of the model's performance is provided. At this point, the cross-linguistic predictions are discussed. Finally, some potential extensions of the model to linguistic problems of interest—polysemy and deixis—are considered.

3. Cross-linguistic Typology

The model clearly satisfies the desideratum of providing a typology of spatial semantic systems. The difficulty is that it is not clear what is contained in this typology. Regier focuses mainly on two properties: (1) an **intermediate sequentiality prediction**, which holds that spatial meanings are encoded in terms of the primitives *source, path,* and *destination* only, so there can be no language that uses two different closed-class forms to distinguish two events that differ only in the sequence of events along the path; (2) an **endpoint configuration prediction**, which says that a language with a word meaning "through" or "out of" must also have a word meaning "in" (in either the static "withinness" sense, or the "into" sense). The first of these predictions follows trivially, as Regier notes, from his stipulation that *source, path,* and *destination* are the primitives that the model must work with. This prediction is a standard type of claim about semantic universals, and as such it deserves some attention, but in the context of a connectionist learning device that is supposed to be making typological predictions, it is a bit of a nonstarter. The question one really wants an answer to is: What constraints, if any, do the connectionist features of the model (the hidden nodes, the gradient descent learning, the distributed representations) put on the class of learnable languages? Similarly, Regier suggests that the endpoint configuration prediction stems

from the assumption that "the model learns to categorize based on only those spatial features that occur at the end of some event it has seen" (p. 157). Again, the prediction seems to stem trivially from one of the features of the encoding system, not from any property of the learning mechanism.

Now, one might argue that the connectionist part of the model is not supposed to do any constraining of the typology, and that its value lies entirely in its role as a learning mechanism, which is, perhaps, justification enough for its inclusion. The trouble is, the learning mechanism probably is putting some typological constraints on the model, but no effort is made to find out what these constraints are. The endpoint configuration issue makes this point clear. Regier's implication that the endpoint configuration prediction is a hard constraint on the model's behavior may not be precisely true. The prediction about "through" stems from the fact that in order to have passed through the landmark, the trajector must have been in the landmark at some earlier time frame; only if the model has recognized "in" in an earlier time frame, Regier implies, will it succeed in storing this fact in one of its maximum-detecting units. Suppose, however, that the random initial weights of the network happened to be set in such a way that a certain node in its hidden layer reached its maximum value whenever the trajector was "in" the landmark. Suppose, furthermore, that the network with this random initial setting was trained on a language that had a word for "through" but no word for "in." It seems probable that such a network would be able to preserve the useful features of its initial weight setting and learn to recognize instances of "through" despite the fact that there is no word for "in," thus contradicting the endpoint configuration claim. However, such a result would contradict the endpoint configuration claim in only a trivial sense because the chances of encountering such a random initial weight setting are probably very slight. What is the moral? Connectionist networks like Regier's probably do have typological implications in virtue of their connectionism. These implications are of a probabilistic nature, not an absolute nature, so the fact that sufficiently large networks can emulate Turing machines (Siegelmann and Sontag 1991) is not directly relevant. The presentation of Regier's model as a typological theory would be of much greater interest if it assessed these implications.

4. Within-Language Generalization

The treatment of learning in the absence of negative evidence, although not much more analytical than the treatment of typological constraints, is nevertheless much more appealing because it points the way to a more in-depth treatment. Regier notes that a principle of **mutual exclusivity** has been posited in several guises in the child-language literature (pp. 62–3): a child prefers that each entity have only one name. Clearly this cannot be an absolute constraint, since many entities have multiple names. (Likewise, many spatial relationship situations can be described by several spatial markers—for example, a trajector can be simultaneously *outside of* and *above* a landmark.) Regier translates this observation into a learning system in which the positive occurrence of a word in a situation provides a strong error signal to a network learning to use the word, but the failure of the occurrence of that word provides only a weak signal. Regier provides evidence that it is only a particular relationship between the strengths of errors on occurrences and non-occurrences that produces the right generalization performance. He finds, for example, that if this parameter is set at its highest value (so that non-occurrences generate no error signal at all), then there is rampant overgeneralization: every position is deemed to be "outside" the landmark. On the other hand, if the value is set too low (so that non-occurrences count quite strongly), then

only a proper subset of the actual "outside" instances is recognized. Only when the relationship has an appropriate intermediate value does the network actually recognize as instances of "outside" precisely those positions that are not contained in the landmark. What's nice about this result is that the domain is one in which we are able to say in topologically precise terms what "correct generalization" is. It may even be possible to formulate principles in terms of categorical notions like inclusion, contact, directional orientation, and so on, which will tell how to generalize from given sets of instances. One could then, perhaps, derive the correct ratio of error strengths, rather than having to stipulate it. In this regard, it might be of some interest to try to incorporate Regier's results into the developing theory of the relationship between cost functions and generalization performance (Rumelhart et al. 1995, Vapnik 1995). The payoff of such an endeavor, if it succeeded, would be an analysis of how topologically precise principles of generalization can arise from a statistical approximation mechanism. Such a result would go some way toward eliminating the haze surrounding the interpretation of neural networks as scientific theories. Regier's book does a good job of laying the groundwork for such an investigation.

5. Readership Recommendations

As the previous section has indicated, the book may be of particular interest to mathematically minded learning theorists who are looking for a domain in which to probe the challenging problem of generalization. On the other hand, it is also a good introduction to connectionist models for anyone who does not have a mathematical background. It should be read by cognitive linguists in general for its effective employment of the hypothesis that a linguistic theory must be simultaneously a theory of language and a theory of perception. It will be of interest to lexical semanticists for its treatment of spatial markers and polysemy. Although its treatment of learned structural constraints, as opposed to pre-encoded structural constraints, leaves something to be desired, as I have noted, the issue and implementation are laid out with appealing clarity so that anyone interested in this dichotomy should take a look at it. Finally, the book should be of interest to anyone who believes in the value of trying to weave together constraints from a variety of domains into one large computer simulation, as well as anyone who particularly doesn't.

Rumelhart, David E., Richard Durbin, Richard Golden, and Yves Chauvin. 1995. Backpropagation: The basic theory. In Yves Chauvin and David E. Rumelhart, editors, *Backpropagation: Theory, Architectures, and Applications.* Lawrence Erlbaum Associates, Mahwah, NJ, pages 1–34.

Siegelmann, Hava T. and Eduardo D. Sontag. 1991. Turing computability with neural nets. *Applied Mathematics Letters*, 4(6):77–80.
Vapnik, V. 1995. *The Nature of Statistical Learning Theory.* Springer, New York.

Whitney Tabor, a postdoctoral fellow in the Mind Articulation Project at MIT, has developed neural network models of language change and language processing. Tabor's address is MIT, Bldg. 20C-228, Cambridge, MA 02139; e-mail: tabor@mit.edu