Constrained connectionism and the limits of human semantics: A review essay ...
Robert M French
Philosophical Psychology; Dec 1999; 12, 4; Academic Research Library
pg. 515

PHILOSOPHICAL PSYCHOLOGY, VOL. 12, NO. 4, 1999

Constrained connectionism and the limits of human semantics: a review essay of Terry Regier's *The human semantic potential*

ROBERT M. FRENCH

ABSTRACT *Taking to heart Massaro's [(1988) Some criticisms of connectionist models of human performance,* Journal of Memory and Language, 27, *213–234] criticism that multi-layer perceptrons are not appropriate for modeling human cognition because they are too powerful (i.e. they can simulate just about anything, which gives them little explanatory power), Regier develops the notion of constrained connectionism. The model that he discusses is a distributed network but with numerous constraints added that are (more or less) motivated by real psychophysical and neurophysical constraints. His model learns "static" prepositions of spatial location such as* in, above, to the left of, to the right of, under, *etc., as well as "dynamic" prepositions such as* through *and the Russian* iz-pod, *meaning "out from under." The network learns these prepositions by viewing a number of examples of them. Very importantly, this book tackles—and goes a long way towards resolving—the problem of the lack of negative exemplars (i.e. we are only very rarely told when something is* not above *something else), which should lead to overgeneralization, but does not. This book is a significant contribution to connectionist literature.*

It is probably not appropriate for an academic review of a technical book to read like an exclamation point filled review of a new Hollywood movie, but, on the other hand, I don't want to tone down my enthusiasm for Terry Regier's exceptional book, *The human semantic potential: spatial language and constrained connectionism.* It is everything a good book on connectionist modeling should be. So, if you do connectionist modeling, or any kind of modeling, for that matter, your personal library will not be complete without this book. Buy it, as they say in the trade press, and get a second copy for a friend.

If nothing else, *The human semantic potential* is a masterpiece of clarity. Over and over I found myself raising this or that objection to some point in the text and, almost without fail, the objection was subsequently presented—and usually answered—several pages later. The book is organized in such a way that the reader

Robert M. French, Quantitative Psychology and Cognitive Science, University de Liege, Liege, Belgium. Email: rfrench@ulg.ac.be

0951-5089/99/040515–09 © 1999 Taylor & Francis Ltd

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

moves smoothly from overarching issues to particulars, from explanatory examples to the actual working of his system. And this is no easy task, however effortless the author makes it all seem.

Regier tells us right from the beginning what problems he intends to tackle and why they are important. And then, when he has shown us how everything works, he sums it all up again, elegantly and clearly, as follows:

> The primary scientific thrust of this work has been to characterize the human semantic potential, that capacity for meaning that is shared by all humans and is shaped through the process of language acquisition into the semantic system of the speaker's native language. The idea has been to determine what limits there are, if any, to this capacity. Can any word mean anything, or are there clearly specifiable constraints on the possible semantic contents of words? This core question has guided and informed all of the work described here. (p. 186)

So, what has Regier done? He has created a connectionist model of understanding various "closed-class" sets of linguistic items. These are words, such as prepositions, which admit few new members and are relatively few in number. In particular, his model learns "static" items such as *above*, *below*, *to the left of*, *to the right of*, *inside*, *outside*, *on*, and *off*, and dynamic items, such as *through* and *iz-pod* (a Russian preposition meaning "out from underneath"). George Lakoff (1987) and Lakoff and Johnson (1980) are, in many ways, the spiritual forebears of many of the ideas underlying Regier's model, in particular, the notion that "space serves as a fundamental conceptual structuring device in language" (p. 19). The brand of connectionism Regier chose to use to model closed-class lexeme acquisition is also indebted to Jerome Feldman's so-called structured connectionist models (Feldman, 1989; Feldman *et al.*, 1988). These are essentially networks whose nodes have symbolic interpretations (e.g. "dog" might be a node in such a network) and whose architecture specifically reflects various cognitive structures. Regier constrains his network by explicitly incorporating "a number of structural devices motivated by neurobiological and psychophysical evidence concerning the human visual system" (p. 2). The network learns closed-class linguistic terms by observing numerous examples of the concept and generalizing from them. (Figures 1 and 2, for example, show "movies" that could have been used to train the network on the preposition *through*.)

The goals of Regier's book are clearly laid out from the beginning. He wants to characterize systems that could adapt themselves to the various structuring of space manifested in the world's languages. He explores how such systems could be made to learn and generalize without the benefit of negative evidence. The problem is a fundamental one in child language acquisition: even though children might be explicitly told when something is *above* a box, they are only very rarely told when something is *not above* the box. Under these circumstances, why do they not radically overgeneralize when using the word *above*? Finally, he wants to know what this model might tell us about possible semantic universals and about the human semantic potential. In short, are our capacities to classify essentially limitless, allowing any sort of spatial relation or event whatsoever to acquire its own name?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Constrained connectionism

The problem that Regier's model attempts to solve, i.e. the learning of closed-class lexical items, is more than anything else, a vehicle for defending a particular philosophy of connectionist modeling. While a standard argument from the traditional artificial intelligence camp (see, for example, Fodor & Pylyshyn, 1988) focuses on what connectionist models cannot do, Massaro (1988) took a wholly different approach. He claimed that they were inadequate not because they were too weak, but, rather, because they were too powerful to be scientifically meaningful. In other words, since the computational power of connectionist models appears to be more or less unbounded, they lose explanatory power. As Regier points out, if connectionist networks, "can give rise to both human and nonhuman behavior, they make poor models of human learning" (p. 10).

Regier takes Massaro's criticism seriously and is the basic reason for his development of what he calls *constrained connectionism*. He introduces neurobiologically and psychophysically plausible constraints to his model. He takes great pains to explain the importance of *independently motivated* constraints that are unrelated to the problem to be solved. This is of utmost importance, since, if the constraints are not motivated in a manner that is truly independent of the problem to be solved, we are back to the problem that plagued traditional AI—namely, hand-coded representations designed to solve a specific problem.

So, are the constraints that Regier applies to his connectionist model plausible, independent of the problem to be solved? Sometimes they are; sometimes they strike me as a bit too motivated by the solution to a particular problem for my liking. Chapter 5 is devoted entirely to explaining the structures that he uses to constrain his model. For example, I am completely comfortable with his use of "filling in" via spreading activation (the idea is that the brain automatically fills in areas by a means of spreading activation, thus allowing it to determine "interior" and "exterior"). I am less comfortable with the author's discussion of orientation combination. Essentially, he gives a number of rules of proximal orientation and center-of-mass orientation that influence our judgment of, say, the term *above*. The problem with this type of structure is that it looks suspiciously like one that is not independently motivated, but rather is incorporated precisely to solve the problem of learning the word *above*. [I have recently learned (Regier, personal communication) that his idea of proximal and center-of-mass orientation is not only poorly motivated, but has been shown empirically to be incorrect.] In addition, as we will see later (Figures 3 and 4), the semantics of the objects to which these terms apply *does* play a role in our understanding of them, even though one has the impression in reading Regier's book that this is not the case.

I would have liked to have seen Regier at least discuss the extent to which various structures could have been the product of learning and which were, on the contrary, more likely to have been hard-wired. Which of his constraints might have emerged if the network had been allowed to learn on a far wider variety of training exemplars? There is no real discussion of this in the book.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Equally problematic is the use of "path buffers" as one of the constraints on his model. Here is a structure that looks suspiciously like it was specially chosen to get the network to learn *through* correctly. The idea is that when looking at a "movie" of one object passing through another, we do not keep track of the actual timing of the events that occurred over the path of movement, but only of the global sequence of events that occurred over the path as a whole. Consider the two scenarios in Figure 1 Both figures are excellent examples of *through*. But what information, exactly, is contained in the buffer? In the first example, we can either view this as an example of *through–through–through* or, simply, *through*. And, depending on context, either interpretation could be valid. What information is kept track of in the path buffer? Is the information in the path buffer different for the two examples?

Now consider Figure 2. He tells us that no language that would distinguish between the two examples of *through* shown because the moment at which the object passes through the object is unimportant. But this is rather strange because, as he points out elsewhere in his book, *through* can be decomposed into a sequence of primitive static concepts: "*outside–inside–outside.*" But let's describe the path of the trajector another way, a way which strikes me as providing an even better description of the actual situations: In the left-hand case: "*outside–inside–(outside* AND *above)*" and, in the right-hand case: "*(outside* AND *above)–inside–outside.*" These are clearly two distinct sequences of primitives making up the path. It is not just a matter of the precise moment at which the events comprising the sequence occur, rather, the order of the two sequences is qualitatively different. I see no *a priori* reason why there could not be a language that could group these two distinct ordered sets of primitives and give each a distinct name.

In summary, I would have liked to have seen a more thorough explanation of whether the constraining structures of the model were, in the author's opinion, innate features or higher level constraints that might reasonably have been learned by an initially unstructured system. Regier anticipates this remark by saying that he

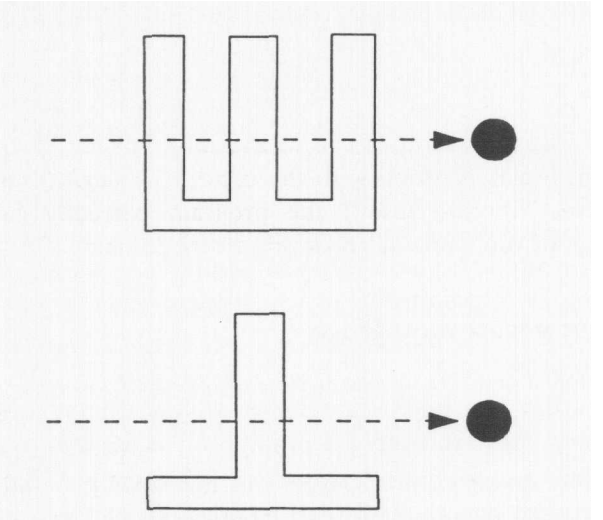

FIG.1. Two different examples of *through*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

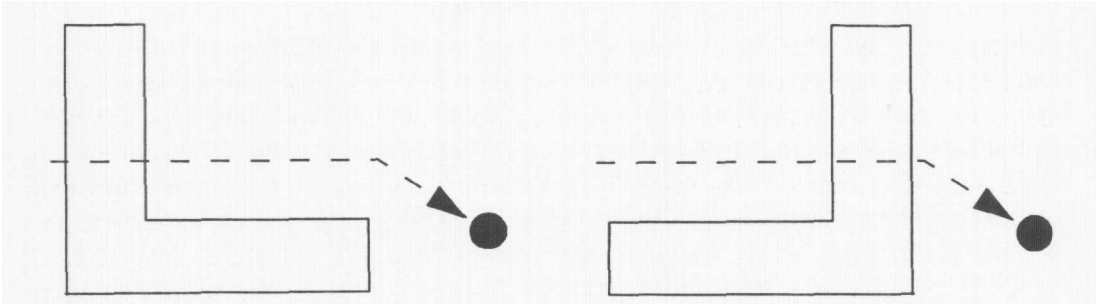

FIG. 2. Regier's model predicts that in no languages will these two paths be described differently.

realizes that the constraints he has applied do not constitute "a complete reductionist grounding" for spatial semantics. He goes on to say, "I simply do not believe that such a thorough reduction is feasible at this time, although I would be delighted to be proven wrong."

Polysemy

"Polysemy" refers to words that have two or more distinct but related meanings. One particularly elegant demonstration of his model was its ability to learn the two meanings of the preposition *in*. These meanings are generally confounded in normal English speech and can be understood by considering the following two sentences: "He walked *in* the room" and "He was standing *in* the room." I vividly remember an explanation of this difference from a junior high school grammar book. In the first case, one had to use "into" (although no one ever did) and, in the latter case, "in." To illustrate the difference, the sentence, "John burst into the room," was accompanied by a drawing of John opening a door and racing into a room, while the sentence, "John burst in the room," showed John exploding in the middle of the room! When Regier's system learns a series of ten examples of *in*, some of which involved "motion into" and the others simply involved the notion "inside," it does not distinguish the two meanings. (This is shown by a cluster analysis of the hidden unit activation patterns of the exemplars of *in* meaning "motion into" compared to those associated with the exemplars of *in* meaning "inside.") On the other hand, when the program learns *in* along with the concept *through* (one of whose component parts involves "motion into") the program correctly categorizes the two different meanings of the word *in* according to its context.

Learning without negative evidence

Perhaps the most outstanding contribution of Regier's modeling effort is to have clearly shown how learning of closed-form lexemes might take place in the absence of negative evidence. The author devotes an entire chapter to this central question. The question to be answered is: "[H]ow can the child generalize from the input without *overgeneralizing* to include inappropriate usages, if these usages have never been explicitly flagged as infelicitous?" (p. 59). The book is worth reading for this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chapter alone. He implements a solution based on the so-called *principle* (actually, heuristic) *of mutual exclusivity* (Markman, 1987) which supposes that children make the assumption that any given object may have only one name. When this heuristic is applied to the domain of spatial terms it works relatively well, if not perfectly, since a positive instance of one concept usually is a pretty good implicit negative instance for all others. For example, a positive instance of *above* is a good negative instance of *below, inside, to the left of, to the right of, etc.* Of course, it is not a good negative instance of *outside*, since something can be both "above" and "outside." But Regier demonstrates that the fact that some pairs of items are not mutually exclusive is not a serious problem. In short, as long as there are a sufficient number of examples, the mutual exclusivity heuristic is good enough to learn the concepts correctly. Crucially, the weight attached to implicit negative evidence is less than for positive evidence. In the chapter devoted to learning without explicit negative evidence, Regier takes us step by step through the problem and his analysis of it. He shows us how he implemented this heuristic in his model (including possible variations of the heuristic) and demonstrates how the model learns closed-class lexemes without it (terribly) and with it (well). There are many examples to accompany the details and the chapter is written with unrivalled clarity.

The closed-class lexeme "micro-domain"

One of the most encouraging aspects of Regier's work is his return to the exploration of a "micro-domain," although he never explicitly calls it that in his book. One difficulty with using micro-domains is that they have to be restricted enough to be able to be studied, but rich enough to allow us to learn something from them. The micro-domain of closed-class lexemes is very rich, indeed, and is exactly the place that one should start the kind of language acquisition modeling Regier is proposing. But one of the questions that I would have liked to see Regier discuss, if not implement, would have been the extent to which he is proposing *general* language learning mechanisms. In other words, to what extent does his model potentially scale up to a larger domain that would include, say, verbs of motion? Unfortunately, he hardly addresses this issue. Does he believe that the constraints that he has placed on his model, inspired by biological and psychophysical mechanisms, would suffice for a more complete model? Or is much more needed in the way of constraints? He does not tell us. The extensions that he discusses are, however interesting, relatively limited in scope. I am not suggesting that Regier should have attempted to implement a larger model, but it certainly would have been nice had he discussed in more detail where we might go with the present model and what other constraints might reasonably be expected to be added to the model when it is scaled up to broader classes of items.

The need for semantics

In addition, there are many examples of the use of these closed-class lexical items that the model would have difficulties learning because semantics is entirely lacking in this model. For example, consider the following drawing of a bowl of fruit (Figure 3) filled with apples. For the figure on the left, we would say, without hesitation,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

FIG. 3. The apple on the left is *in* the fruit bowl; on left it is *above* it. But, in fact, they are the same height above the bowl in both cases.

"The red apple is *in* the fruit bowl," whereas for the figure on the right, even though the apple is in exactly the same position with respect to the fruit bowl, we would say, "The apple is *above* the fruit bowl," a preposition that it would not even occur to us to use when the fruit bowl was full. And yet, even if we consider a convex hull definition of what would constitute the "inside" of the fruit bowl, we cannot capture the former sense of *in*. (Thanks to Stephanie Kelter for this example.)

Similarly, a picture like the one in Figure 4, even if it fit the perfect convex hull definition of *in* would not be a good example of "the apple is *in* the fruit bowl." Why? Because we have semantic knowledge about the proper orientation of fruit bowls and this knowledge radically affects our perception of the concept "in." On the other hand, let's say I describe the identical form in Figure 4 as an igloo and replace "apple" by "Eskimo." All of a sudden, it becomes a perfectly good example of *in*.

Now, of course, Regier could come up with specific rules that would allow his system, in a post hoc manner, to resolve each of these particular problems. But then, of course, the problem with this is obvious. The constraints placed on the system are no longer independently motivated, but rather are motivated by the specific problem.

The point here is a general one. In some sense, Regier's model is attempting to learn the prepositions of spatial semantics in a context-independent manner. There is no *a priori* reason to assume that the program would not consider both cases in Figure 3 to be equally good examples of *above* and equally bad examples of *in*. And it would presumably find that Figure 4 is an excellent example of *in*, which, *depending on the semantics of the situation*, may or may not be the case. The point is that our semantic knowledge about fruit bowls and apples, igloos and Eskimos, comes into play here and provides a context that changes the spatial semantics of the preposition *in*. But this knowledge is nowhere present in Regier's program, and it seems to work so well that we all but forget just how important this world-knowledge is. The fruit bowl example is not some trick-question exception either. Closed-form lexemes, just like concepts like "dog" and "run," are part of real language and real

FIG. 4. A bad instance of *in* if the form is a fruit bowl and the object is an apple; a good example if the form is an igloo and the object is an Eskimo.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

language is grounded in semantics. And real semantics requires knowledge about fruit bowls and piles of fruit, about igloos and Eskimos, etc. But the skeletal "movies" that Regier uses are devoid of real-world semantics. Consequently, we tend to forget that we are dealing with cases where contextual semantics is important. Ultimately, if we are ever to hope to resolve the difficulties of computer understanding of everyday language, our knowledge about the world and the relationships between the objects in the world, etc., *must* be incorporated into our program.

Minor problems

There are, inevitably, a few minor problems with the book. For example, Regier's editors really should have changed the book's stuffy title, which is perhaps fine for a black-bound doctoral dissertation, mostly designed to allow parents to prove to the neighbors that their kid really was the smartest on the block, but less acceptable for a work that people are actually going to read and refer to. And, most seriously, the index of this dense, 200 + page book consists of less than two anemic pages of entries. This is completely unacceptable and limits the book's quality as a reference document. Why MIT Press would have accepted such an impoverished index is beyond me.

Conclusion

In conclusion, Terry Regier has written a deeply thoughtful book about modeling certain aspects of language acquisition. It stands as a model of how any good book on computational modeling should be written. The author goes to great lengths to make everything crystal clear. In fact, his writing is sometimes so limpid that one has a tendency to forget just how hard a problem he tackled. One finds in this book exactly the right mixture of general discussion, clear explanation and elegant computational modeling. Many researchers tackle hard problems and the best of them actually make some headway in solving them. But rare, indeed, are those who can undertake daunting problems, make significant progress towards their solution, and then tell us about what they did clearly, persuasively and with deep insight. That is precisely what Terry Regier has achieved in his marvelous book, *The human semantic potential.*

References

FELDMAN, J. (1989). Neural representation of conceptual knowledge. In L. NADEL, P. CULICOVER & R.M. HARNISH (Eds) *Neural connections, mental computation.* Cambridge, MA: MIT Press.

FELDMAN, J., FANTY, M. & GODDARD, N. (1988). Computing with structured neural networks. *IEEE Computer, 21,* 91–104.

FODOR, J. & PYLYSHYN, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition, 28,* 3–71.

LAKOFF, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind.* Chicago: University of Chicago Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LAKOFF, G. & JOHNSON, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

MARKMAN, E. (1987). How children constrain the possible meanings of words. In E. NEISSER (Ed.) *Concepts and conceptual development: ecological and intellectual factors in categorization.* Cambridge: Cambridge University Press.

MASSARO, D. (1988). Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 213–234.

REGIER, T. (1996). *The human semantic potential: spatial language and constrained connectionism.* Cambridge, MA: MIT Press.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.