The Proper Treatment of Semantic Systematicity

Robert F. Hadley (hadley@cs.sfu.ca)
School of Computing Science and Cognitive Science Program
Simon Fraser University
Burnaby, B.C., V5A 1S6

Abstract

Connectionist-minded philosophers, including Clark and van Gelder, have espoused the merits of viewing hidden-layer, context-sensitive representations as possessing semantic content, where this content is partially revealed via the representations' position in vector space. In recent work, Bodén and Niklasson have incorporated a variant of this view within their conception of *semantic systematicity*. Moreover, Bodén and Niklasson contend that they have produced experimental results which not only satisfy a kind of *context-based, semantic systematicity*, but which, to the degree that reality permits, effectively deals with challenges posed by Fodor and Pylyshyn (1988), and Hadley (1994a). This paper examines the claims of Bodén and Niklasson. It is argued that their case fatally involves a fallacy of equivocation. In addition, it is argued that their ultimate construal of *context sensitive semantics* employs lax, incorrect standards.

Introduction

The expressions 'Strong Systematicity' (SS) and 'Strong Semantic Systematicity' (SSS) were introduced and formally defined in Hadley (1994a) and Hadley (1994b), respectively. These definitions were intended not only to clarify the nature of attempts (recent at that time) to satisfy Fodor's and Pylyshyn's (1998) well known systematicity challenge, but to highlight the remaining distance between what those attempts had accomplished and what Fodor and Pylyshyn were demanding. Detailed explanations of SS and SSS are provided in section 2, but for the present the following (oversimplified) characterizations should suffice. A connectionist network (or human agent) exhibits SS provided it learns to generalize a *significant fraction* of its vocabulary to novel syntactic positions within both simple and complex sentences. In contrast, an agent or network satisfies SSS provided it not only exhibits SS, but can assign correct meaning representations to any of the novel test sentences which could be used to establish the presence of SS in that agent.

Now, Niklasson and van Gelder (1994) presented connectionist experiments which, in their view, satisfied at least the conditions required by SS. My (1994b) reply to their article described their results as a "borderline case" of SS, and I pointed out several difficulties with their work. These difficulties included: (a) only a single novel term was ever employed, and (b) the encoding of this novel term had been carefully crafted to ensure that its eventual vector representation would fall exactly in the centre of the vector space that kindred, non-novel terms occupied. These two defects (as I viewed them) contributed very substantially to the network's "success" at generalization.

In recent work, Bodén and Niklasson (2000) present experiments and arguments designed to show that connectionist networks can not only satisfy SS, but also *a kind of* strong semantic systematicity, at least when the concept of `semantic representation' is construed in a fashion which they believe is fair to connectionism. Indeed, they forthrightly claim that "In the experimental section we shall show how the proposed architecture is an extension of the work of Niklasson and van Gelder, intended to remove Hadley's reservation" (p. 129). They also state that "... we contend that the connectionist metaphor is not only leveling with Fodor and Pylyshyn's (1988) challenge but also with Hadley's (1994a, b) revised challenge of semantic systematicity" (p. 139) and further, "The connectionist system we present in the following will be able to assign relevant semantic content to novel tokens appearing in test sentences which could demonstrate strong systematicity." (Bodén and Niklasson, 2000, p. 113).

In what follows, I will argue that B&N have not produced, in their present (2000) work, any convincing example of a network's displaying SS, much less a kind of SSS. As they acknowledge, only one of their experiments is even intended to avoid my 1994 criticism. I contend that this one crucial experiment is fatally flawed, because B&N fail to show that their network ever successfully processes a 'novel test sentence'. Rather, B&N fall prey to the fallacy of equivocation and employ the expression 'novel test sentence' in an unusual, and implausible fashion. Moreover, I argue that they adopt mistakenly lax standards for what constitutes a *correct semantic representation*. Significantly, my examination of this latter issue has relevance beyond B&N's work. For, B&N's remarks on semantics echo similar comments and confusions found in a worrisome range of connectionist publications and conference discussions.

Learning-Based Definitions of Systematicity

In Hadley (1994a), a hierarchy of degrees of learning-based systematicity was introduced. For purposes of the discussion which follows, it will be helpful to have in mind a brief paraphrase of a *portion* of this hierarchy.

- Weak Systematicity. An agent is at most weakly systematic if, after training, it can process "test" sentences (or symbol sequences) containing novel combinations of words (symbols), *but cannot* process sentences containing familiar words in positions which are novel to those words.
- An agent is strongly systematic (SS) if and only if "it can correctly process a variety of novel *simple* sentences and novel *embedded* sentences containing previously learned words in positions where they *do not appear* in the training corpus (i.e., the word within the novel sentence does *not appear in that same syntactic position* within any *simple or embedded* sentence in the training corpus) ... Also, ... training corpora which are used to induce strong systematicity must not present the entire training vocabulary in all the legal syntactic positions, but should refrain from doing so for a *significant fraction* of that vocabulary." (Hadley, 1994a, pp. 250-251).

The forms of systematicity just listed do not require that an agent be capable of semantically interpreting the sentences in question. However, F&P's examples of systematicity included cases where an agent assigns meaning to the sentences involved (e.g., whoever *understands* 'John loves Mary' can also understand 'Mary loves John'). For this reason, a more demanding criterion (given below) of systematicity was offered in (Hadley, 1994b).

- Strong Semantic Systematicity (SSS) "A system possesses semantic systematicity if it is strongly systematic and it assigns appropriate meanings to all words occurring in novel test sentences which (would or could) demonstrate the strong systematicity of the network" (Hadley, 1994b).

Now, it must be noted that B&N do not claim to satisfy the precise details of my definition of SSS. At one point they say,

"The results presented herein do not achieve exactly what semantic systematicity requires. Instead, we have shown that by redefining some central concepts in folk psychology in terms of connectionist primitives a similar kind of context-based systematicity can be achieved. In the following, we shall still use Hadley's levels (weak, quasi and strong) of systematicity to qualify what has been achieved. "

Nevertheless, in the passages quoted earlier, especially when they say "The connectionist system we present in the following will be able to assign relevant semantic content to novel tokens appearing in test sentences which could demonstrate strong systematicity", they strongly imply that they have very nearly satisfied the requirements for SSS, their caveat being that they employ a conception of 'semantic content' which they believe to be most suitable to connectionist research. Moreover, B&N state, in effect, that they will and have produced an experimental result which lays to rest my published 1994 qualms about Niklasson's and van Gelder's 1994 work. For these reasons, I wish to emphasize certain crucial aspects of my definitions of SS and SSS. In particular, both SS and SSS require (i) that previously known words be used in novel positions within (post-training) test sentences; (ii) a significant fraction of the vocabulary of the training corpus must be presented in these novel positions; (iii) the 'novel positions' in question must appear in both simple sentences and embedded clauses.

Of all experiments described in their (2000) paper, not one satisfies points (i) and (ii) above. In addition, as will emerge, their crucial (*coup de grace*) experiment entirely ignores condition (iii). In light of these points alone, the passages I have quoted from B&N seem at least misleading.

Presently, we shall consider the view of "semantic content" that B&N put forward, as they set the stage for the experiment they believe to have attained *a kind of* SSS. Before examining details, however, I would ask the reader to note that nothing in my definition of SSS assumes a classically based semantic theory. My definition only requires that the agent "assigns appropriate meanings to all words occurring in novel test sentences which (would or could) demonstrate the strong systematicity of the network". I have left it an open question how appropriate meanings are represented.

It is important to bear in mind, however, that both my definition of SSS and Fodor and Pylyshyn's original characterization of systematicity were introduced in the context of examples of sentences found in natural language. In both cases, the terms `semantics' and `meaning' were used as they are commonly understood by philosophers and linguists. In particular, the semantics (or meaning) of a declarative sentence in natural language was assumed to be intrinsically connected to the ability of such a sentence to describe or express external situations (or states of affairs) which could *render a given sentence true*.

Since this is so, within the definition of SSS, the phrase "assign appropriate meanings to all words occurring in novel test sentences ..." must be understood against a background of *standards of correctness*. Any purported demonstration that SSS has been attained (or even nearly attained) by a network must present convincing evidence that the "novel test sentence" has been assigned a semantic representation that is semantically coherent and correct.

Bodén and Niklasson's Treatment of "Semantic Content"

B&N are much concerned that "semantic content" be understood in a (theory-laden) fashion that, in their view, does justice to the underlying assumptions of non-classically-oriented connectionism. For this reason, they stress the need to realize that semantics, properly understood, deals with context-sensitive, distributed representations. Nevertheless, their initial characterization of semantic content is certainly compatible with the notion that the semantics of a sentence concerns the relation between the sentence and a possible state of affairs which could render the sentence true. They say, for example, "... the focus in semantic systematicity is on the meaning or content of representational tokens (i.e., what they refer to in the represented world)." The view that semantic content concerns the ability of tokens in a sentence to *refer to* items in a represented world harmonizes nicely with the philosophical-linguistic conception of semantics described in the preceding section. Indeed, it is only because some of the tokens in a sentence refer (or potentially refer) to a represented world that a declarative sentence can have truth conditions.

Thus far, I have no quarrel with B&N's view of semantic content. It is essential to realize, however, that the ability of tokens to *refer to* objects in a represented world places strong constraints on the degree of context sensitivity of the meaning of words in sentence. The word 'rabbits' denotes exactly the same class of objects in each of the following sentences: "Ferraris are faster than rabbits", "Rabbits are faster than turtles". As Fodor and Pylyshyn (1988) correctly remind us, it is only because of this consistency in reference (or meaning) that any conclusion logically follows from those two sentences. B&N effectively acknowledge this point, when they cite Clark (1993) (see their p. 116), but they do not further discuss the conflict between their emphasis on context sensitivity and the constraints just described. This is unfortunate because the view of semantic content that they proceed to present largely abandons the *essential* idea that semantic content pertains to the ability of "tokens" and sentences to refer to aspects of an externally represented world. As their paper unfolds, B&N reveal that, in their view, the semantic content of an internal representation is determined by three factors, namely, (i) the word order (syntactic) constraints imposed by the input training data, (ii) the position that a representation's activation pattern occupies in vector space, and (iii) the various associations that the representation acquires during training.

Now, it is crucial to realize that factors (i) and (ii) could not be a sufficient source of semantic content for words or sentences. This follows from the following fact. It is possible (and indeed this sometimes occurs) to create training corpora on the basis of artificial grammars and vocabularies which have no prior semantic content whatsoever. The sentences within these corpora incorporate word-order (syntactic) constraints imposed by the artificial grammar in question, and these constraints may be quite elaborate. Nevertheless, the sentences within the corpora simply do not possess any descriptive (or referential, or semantic) relationship to an external world, state of affairs, or situation. In such a case, internal distributed activation patterns (which develop on hidden layers of a network trained on the corpus) cannot be representing the semantic content of such sentences, because the sentences are utterly meaningless -- they *refer to nothing*. Consequently, the positions such activation patterns occupy in vector space cannot be contributing to the semantic content of those sentences.

It is relevant to note, moreover, that even if sentences in the input corpus are selected from a natural human language, and so presumably possess meaning, any network architecture and training regime that generates hidden layer (HL) activation patterns *merely on the basis of the contextual constraints within the training corpus* would be generating HL patterns of precisely the same kind as are generated for the utterly meaningless sentences described above. That is, the HL activation patterns in each case would merely comprise statistical information about co-occurrence patterns among symbol tokens. Information of this type cannot constitute semantic information, because this type of information is identical in structure both when the input corpus contains only meaningless sentences, and when it contains sentences known to have meaning.

We have now seen that factors (i) and (ii) are not sufficient to endow an internal representation with semantic content. Therefore, if B&N's "working semantic theory" is to be credible, much depends on the plausibility of the supplemental factor (iii), which concerns the *associations* formed by internal representations. Now, the contention that HL activations *acquire* semantic content, in part, from their associations with other data (via the intervening weights) offers promise of providing at least the foundations of an acceptable semantic theory. However, as philosophers of language are well aware, the associative correlations must be of the *right kind* if they are to provide the basis for an adequate account of semantic reference (or denotation). In particular, it must be the case that the associative correlations be rich (or complex) enough to explain how the ``inner representations", produced by the understanding of declarative sentences, could *express* a complete state of affairs (or a situation) in the external world. In order for this to be possible, the inner representations themselves must possess sufficient complexity to permit them to be *mapped* onto external states of affairs. (This is so even if the mapping is performed solely by complex weight vectors within the agent's brain).

Unfortunately, B&N never address the issue of the *kind or complexity of associations* that must be formed if their HL representations are to acquire semantic content. They do not, for example, require that this "associated data" satisfy any standards of "richness" or correctness as one would expect to see when the HL patterns are purported to be representations of the meaning of natural language sentences. Indeed (as we shall soon see), B&N place no constraints whatsoever on the nature of the data that, during training, becomes associated with HL activation patterns.

B&N's Systematicity Experiments

In the "Experiments" section of their paper, B&N describe two types of experiments which they believe to exhibit at least strong systematicity. In what follows, I refer to these as the "Type 1 -- Default-Based Experiments" and the "Type 2 -- Crucial Experiment". Experiments of both types employ distributed representations generated by RAAM networks.

Let us suppose that a given RAAM has input and output layers that each contain two separate regions of "bits". Within the leftmost region, one may present a binary encoding of a given term, say, 'cat'. Within the rightmost region, one may present a binary encoding of a general category that 'cat' belongs to, say, 'noun'. During training of the RAAM, the hidden layer receives information from both the left and right input regions, and over time develops a condensed distributed encoding which blends information from the two input regions. In this way, a distributed encoding for `cat' can be created which contains considerable information about the category or class of that term, 'noun'. As we shall see, B&N employ this kind of "class-based encoding" of terms in their systematicity experiments.

Type 1 -- Default-Based Experiments

B&N acknowledge that their Type 1 experiments do not avoid a criticism which I voiced in my 1994 reply to the claims of Niklasson and van Gelder (1994). (See Hadley, 1994b, for full details.) As noted earlier, my 1994 critique described several problems with Niklasson's approach to encoding and training, but the criticism that B&N currently acknowledge concerns the RAAM-generated distributed encoding assigned to the single (putatively) novel term that Niklasson employed. I had, in 1994, complained that Niklasson had biased his network's results by assigning to the solitary "novel" term (call it NT) a distributed encoding which shared many featural values with all non-novel terms appearing within precisely the same syntactic position as NT occupied in the test sentences.

As noted, B&N recognize that their Type 1 experiments are open to the objection just explained. However, in my view, the current (Type 1 -- default-based) experiments are open to a more severe version of this criticism, for the following reason. The current experiments involves *default reasoning* with terms, such as 'sparrow', 'penguin', 'tweety' and 'ernie' which are assigned to *classes*, such as 'bird'. All such terms are assigned distributed encodings by a RAAM network, whose encoding processes are influenced by error feedback from another task-oriented (transformation) network. RAAM generated encodings for terms such as 'tweety' and 'sparrow' include information about the 'class' that these terms pertain to, e.g., 'bird'. During the encoding process, B&N have ensured that distributed encodings of the two "novel" terms they employ *not only* share a substantial amount (about 50%) of *class-based* featural information with non-novel terms occurring in the identical syntactic position, but these distributed encodings are partially shaped by error feedback derived from *every task* that the "novel" terms are ever involved in. (B&N repeatedly describe the influence of this error feedback, though they find no fault with it.)

Because the latter experiments not only involve the result-biasing technique of pre-assigning class-based representations to putatively novel terms, but involve the task-oriented biasing just described, I submit that B&N have failed to make a credible case for strong systematicity in their Type 1 experiments.

Type 2 -- The Crucial Experiment

As noted earlier, B&N are aware that the experiments discussed above do not meet the published reservations of (Hadley, 1994b). However, they present one final experiment which they believe adequately answers those reservations. While B&N stress that this crucial experiment does *not* satisfy the precise requirements of SSS (as I define it), they do insist that it satisfies *a kind of* strong semantic systematicity. Also, as previously remarked, they contend that the only reason this experiment does not satisfy my SSS, is that they have redefined a number of terms of "folk psychology", and in doing so have adopted a conception of "semantic content" which they believe now provides connectionists with a level playing field.

I have claimed that B&N's conception of semantics suffers from serious difficulties. In what follows, we shall see how these difficulties arise in their crucial experiment. Quite apart from concerns about "semantics", however, their interpretation of this experiment involves a fatal equivocation involving the expression, "novel test sentence". To see this, we must review the general outline of their design.

As in the Type 1 (Default-Based experiments), the Type 2 (Crucial) experiment employs two RAAM networks and a simple, two layer, transformation network. The first RAAM net is used to create class-based distributed representations for the atomic terms. (E.g., the term 'ernie' is given the class-based encoding, 'bird'). A total of three terms ('ernie', 'bo', and 'jack') ever receive class-based encodings during the experiment. During the first of two training phases, class-based encodings are created for 'ernie' and 'bo' on the first RAAM's hidden layer and are extracted for later use in the second RAAM. The third of the atomic names, ('jack') does not receive a class encoding until a second training phase is performed.

The second RAAM, described by B&N as an "assertion encoder", is used to create encodings for four very simple "sentences", namely,

R(ernie fly) [which we may read as "ernie flies"],
R(ernie not-fly), R(bo fly), R(bo not-fly)

It is relevant to note that, although 'bo' has the class-based encoding of 'fish', the assertion encoder RAAM is never trained to generate assertions to the effect that Bo swims or that Ernie (which is a bird) does not swim.

RAAM generated encodings for two of the four sentences shown above are used in the initial training of the last of the three networks, the transformation network. In particular, the transformation network is initially trained to output '1' (or 'true') when the input is R(ernie fly) and to output '0' (or 'false') when the input is R(bo fly). During a later training phase, this same network is trained to output `1' for the assertion R(jack fly).

Now, because 'jack' does not receive a class-based encoding during the initial training phase of the first RAAM, B&N regard it as a novel token. Moreover, the distributed encoding assigned to R(jack fly) was created by simply presenting 'jack' and 'fly' to the two input regions of the second RAAM and extracting the contents of that RAAM's hidden layer. This RAAM received no training on that input during the initial training phase.

Once this second RAAM has created a distributed encoding for R(jack fly), B&N present this encoding as input to the third network (transformation net). At this stage, the transformation network produces no useful response to R(jack fly). Since no class encoding has been assigned to 'jack', this is perfectly understandable. A human would likewise be unable to produce any helpful response to R(jack fly) at this stage, since the human would have no idea whether `jack' is supposed to be a bird, a fish, or even mud.

However, B&N next proceed to train the transformation network, for 1000 epochs, on the assertion encoding for R(jack fly). Backpropagation is employed, and the target output during this second training phase is '1'. During this new training phase, error feedback not only alters the behaviour of the transformation network on R(jack fly), but is conveyed back to the hidden layer of the second RAAM, and thence back to the hidden layer of the first RAAM. The input-to-hidden-layer weights of both these RAAMs are modified, during the 1000 epochs just mentioned, using this error feedback.

As we would expect, this second training phase eventually succeeds in associating R(jack fly), within the transformation network, with an output value of '1'. That network is now able to produce '1' for just two input sentences. Note also that this trained association (i.e., producing an output of '1') is the sum total of "associative content" ever given to the sentence, R(jack fly) or to R(ernie fly). Under these circumstances, and given that error feedback is used during this second training phase to shape the input-to-hidden-layer links of the initial class-based RAAM encoder, it is not surprising that this RAAM develops for `jack' the class-based encoding of 'bird'. Nor is it surprising that the hidden layer encoding eventually assigned to 'jack' lies very close, in vector space, to the hidden layer encoding of 'ernie'. After all, the only other assertion ever trained to produce an output of '1' is R(ernie fly), and 'ernie' has the class-based content, 'bird'.

What is surprising (to my mind), is that B&N believe that the results just described entail that this last experiment displays an important kind of strong semantic systematicity. Indeed, the results just cited are their sole justification for the following claim: "The connectionist system we present in the following will be able to assign relevant semantic content to novel tokens appearing in test sentences which could demonstrate strong systematicity." (Bodén and Niklasson, 2000, p. 113). The textual context and precise wording of this quote make it clear that B&N have my SS in mind when they say 'strong systematicity'. Moreover, their discussion of this last experiment makes it entirely clear that they regard the sentence R(jack fly) as the test sentence which is assigned "relevant semantic content", and they regard 'jack' as the novel term. (For brevity, I shall refer to 'R(jack fly)' as sentence 'S').

Now, as the reader will recall, not only my definitions of SSS, and SS, but even the definition of 'weak systematicity' requires that a trained network be tested upon a "novel test sentence". Moreover, in their original characterization of systematicity, F&P are clearly claiming that humans who have the capacity to understand a sentence such as "Mary sees the kitten" will automatically have the capacity to understand systematically related sentences that they have *never encountered before*. The employment of novel test sentences is therefore an essential component of any counterexample to F&P. Yet, at the (post-training) stage where B&N are able to claim some form of success for their network, it would be bizarre to regard S as a novel test sentence. For, at this final stage, their network has been subjected to considerable training upon S (1000 epochs).

Now, it is beyond dispute that as 'novel token' and 'novel test data' are commonly used by connectionists, 'jack' and S are, at the relevant stage, *not novel test data*. Moreover, it is difficult to believe that B&N are unaware of this common usage. Any charitable reading of B&N must, therefore, assume that B&N are using those phrases in some new and surprising sense. Given this, I can only conclude that B&N have committed a serious instance of the fallacy of equivocation.

In any case, B&N imply, more than once, that this crucial experiment deals satisfactorily with my criticism (Hadley, 1994b) concerning the pre-assignment of class-based encodings employed in Niklasson's and van Gelder's 1994 experiments. Yet, that criticism was set in the context of my definitions of systematicity, which assumed the normal understanding of *novel test data*. Any experiment directed at meeting those qualms must employ this same understanding if equivocation and fallacy are to be avoided. The same holds true of my SSS challenge and of F&P's original (1998) challenge. What then are we to make of B&N's summary remark that “... we contend that the connectionist metaphor is not only leveling with Fodor and Pylyshyn’s (1988) challenge but with Hadley’s (1994a, b) revised challenge of semantic systematicity”? B&N’s crucial experiment does not even satisfy the novelty requirements of my “weak systematicity”. Admittedly, B&N have taken care to state that they “do not achieve exactly what semantic systematicity requires”. However, the quotations given above and hitherto demonstrate that B&N have at various points claimed, implied, and suggested that they have dealt with not only my challenges, but that of F&P.

Apart from the foregoing issue of novelty, there remains B&N's clear claim that their network `will be able to assign relevant semantic content to novel tokens appearing in test sentences which could demonstrate strong systematicity'. We have seen that B&N take `jack' to be the novel token, but what of their contention that `jack' has, in the end, been assigned `relevant semantic content'? B&N's belief is that, upon completion of both training phases, both `jack' and `ernie' have been assigned virtually identical semantic content -- the content being the class `bird'. The case they offer for this belief is that, following all training, the HL vectors for `ernie' and `jack' occupy nearly the same region of vector space. B&N describe this spatial region as the `bird' region, though their justification for this ascription is dubious. For, at the crucial time of testing, there is *no reason* to believe that `ernie' is any longer associated with a *correct* encoding of `bird'. This is because, in order to achieve the results they desired, during the final training phase, B&N forced the input encoding for `bird' to mutate during each of 1000 epochs. At the end of these 1000 epochs, the `bird’ input representation differs substantially from the initial encoding of `bird’. Although the *initial* encoding might be regarded as a correct representation of the concept of a bird, there is no reason to think that the mutated encoding is in any way correct. In any case, as explained earlier, the mere fact that the HL vector of a (purportedly) novel term or sentence lies close in vector space to another vector in no way establishes that the `novel HL vector' represents a *correct or coherent* meaning, as opposed to a garbled and degraded version of the remaining vector. Moreover, we have seen that the mere fact that such vectors occupy positions in vector space does not ensure that these vectors represent *any* meaning. We are only entitled to assume that the vectors represent meanings when they possess *appropriate*, and sufficiently elaborate associations within the cognitive agent. Let us consider therefore whether other possible associations are developed for the HL vector of `jack' within B&N's networks.

Recall that the HL encoding of `jack' is used as an input component in generating a distributed representation for the complete sentence `Jack flies'. This latter representation, in turn, eventually becomes associated with an output value of `1' within the final transformation network. Could this last association, then, provide adequate semantic content, albeit indirectly, for the input token `jack'? Clearly, the answer is no. For, any number of other sentences, such as `Bo swims', and `Ernie does not swim', could be trained to produce `1' in the output layer of the transformation network. It is obvious that these differing sentences have rather different meanings. The plain truth is that the potential for all the distributed encodings of these sentences to generate `1' as output -- reflects a task that is too trivial to demonstrate that each of these encodings already possesses an internal structure that would permit *correct and coherent* associations with *referential content* to be acquired. For this reason, it is implausible that `Jack' acquires correct semantic content via the associations developed for `Jack flies' within the transformation network.

Summary

In the foregoing, we have seen that B&N employ a seriously implausible conception of “novel test data” and, in their experiments, unacceptably weak standards for *semantic content*. Moreover, their crucial experiment employs just a single, purportedly novel term and sentence. Likewise, they have ignored issues pertaining to embedded clauses. Given all this, it appears farfetched for B&N to claim that they have “leveled with” the challenges raised both by F&P and Hadley (1994a, b). It is, indeed, difficult to discern what relevance B&N’s results have to those challenges. This is not to say that those results have no value, but the onus now rests upon B&N to explain the relevance of their experiments to systematicity issues, as ‘systematicity’ is commonly understood.

References

Bodén, M. and Niklasson, L. (2000), Semantic Systematicity and Context in Connectionist Networks, *Connection Science*, 12, pp. 111-142.

Clark, A. (1993), *Associative Engines*, Cambridge, MA: MIT Press.

Fodor, J.A., & Pylyshyn, Z.W. (1988), Connectionism and Cognitive Architecture: A Critical Analysis, *Cognition*, 28, pp. 3-71.

Hadley, R.F. (1994a), Systematicity in Connectionist Language Learning, *Mind and Language*, 9, pp. 247-272.

Hadley, R.F. (1994b), Systematicity Revisited, *Mind and Language*, 9, pp. 431-444.

Niklasson, L.F. and van Gelder, T. (1994), On Being Systematically Connectionist, *Mind and Language*, 9, pp. 288-302.

Phillips, S. (1994), Strong Systematicity within Connectionism: the Tensor-Recurrent Network, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 723-727.