Mind & Language ISSN 0268-1064
Vol. 9 No. 3 September 1994

Article
Systematicity in Connectionist Language Learning

ROBERT F. HADLEY

1. Introduction

It is by now widely acknowledged by cognitively-oriented connectionists that human thought and language production display both *compositionality and systematicity*. In large part, recognition of this fact has been fostered by Fodor and Pylyshyn (1988), who argue that prevailing methods of connectionist representation cannot engender the *combinatorial* syntax and semantics necessitated by compositionality. Further, since combinatorial syntax and semantics *cause those systematic relationships* of thought and language that concern Fodor and Pylyshyn (F&P), the very features of connectionist representations that preclude compositionality also preclude systematicity (or so it is argued). Much of F&P's discussion addresses the question whether connectionists can achieve structure-sensitive processing without, in effect, creating connectionist implementations of classical symbol processing systems. F&P argue for a negative answer.

Since the appearance of F&P's paper, a number of connectionists have produced seeming *counterexamples* to the F-P thesis. In the present work I examine six of these apparent counterexamples, which are due to Elman (1990), St. John and McClelland (1990), Chalmers (1990), McClelland and

---

The author gratefully acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (Grant No. A0899) during the period of this research. I am especially grateful to Steven Pinker for his invaluable remarks on a previous version of this paper. Thanks also to Martin Davies, Allan Bennett-Brown, Tony Plate, and an anonymous reviewer of this journal for their helpful comments. Of course, my indebtedness does not imply agreement on the part of these commentators with all or most of my position. Address for correspondence: School of Computing Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada. Email: hadley@cs.sfu.ca

Kawamoto (1986), Pollack (1990) and Smolensky (1990). I argue that, although several of these networks *may* constitute genuine counter-examples to F&P's position, there exists a related, learning-based form of systematicity which connectionist systems (c-nets) have not attained. In what follows I distinguish three degrees of systematicity, ranging from weak to strong. While humans exhibit the strongest of these, I argue that existing connectionist methods have thus far attained, at best, a quasi-systematicity which is still significantly different from human systematicity.

Now, I should emphasize that not all the connectionists discussed here claim to be addressing F&P's thesis, and none of these authors present their results as anything more than suggestive of what occurs in human cognition. Nevertheless, cognitive researchers will naturally wonder not only whether connectionism can meet the challenge of strong systematicity, but whether it can do so under *cognitively plausible* conditions. Although a principled answer to the latter question is beyond the scope of this paper, we may gain insight into the problems involved by considering the plausibility of methods that connectionists have employed *even to achieve weak forms of systematicity*. For this reason, I shall sometimes comment upon aspects of the training methods employed that seem seriously unrealistic.

2. Systematicity

2.1 Fodor and Pylyshyn's Conception

It is widely acknowledged that, to an impressive degree, human language possesses a combinatorial syntax and semantics. One causal consequence of these combinatorial properties is that certain systematic relationships occur in language. A central tenet of F&P's position is that humans exploit these systematic relationships in language comprehension and production. F&P argue that these systematic relationships occur not only in language but in thought—that is, systematicity applies to both external and internal representations. In their view, the ability to exploit systematic relationships is *intrinsic* to human cognition. One such relationship is the interchangeability of various arguments to particular verbs. As F&P remind us, you don't find cognitive agents who can understand (or think) 'John loves Mary', but not 'Mary loves John'.

Now, the interchangeability of a verb's arguments is not as straightforward as one might initially suppose. It is widely accepted among linguists that verbs place differing constraints on their arguments, and these constraints *must be learned* (cf. Pinker, 1989). Furthermore, some verbs are always transitive, others are always intransitive, and others permit direct objects in some contexts but not all. Transitive and intransitive uses of the same verb must both be learned. Thus, a child may have learned to use sentences similar to 'Flowers grow' but not 'Pam grows flowers'. If systematicity involves, *inter alia*, the ability correctly to use a verb's arguments in differing positions, then systematicity presupposes a substantial degree of grammatical competence.

In their discussion of systematicity, F&P assume the grammatical competence of cognitive agents, but they are not concerned with the ontogeny of systematicity. Nor do they discuss possible *degrees* of systematicity. Rather, their focus is upon a particular causal precondition of systematicity, viz., *structure-sensitive* operations. As a consequence, some connectionists (Chalmers, Pollack, Smolensky) have stressed the ability of connectionist networks (c-nets) to perform structure-sensitive operations.

2.2 *Systematicity—A New Formulation*

An important aspect of systematicity, as conceived by F&P, is the ability to *generalize*. The reason is readily apparent: there is a clear sense in which an agent who has never seen 'John' used in the patient role, but is able to understand 'Mary loves John', has generalized. As with F&P, the conception of systematicity presented below also involves a capacity for generalization. However, unlike F&P's conception, the concept of systematicity defined here fundamentally involves issues of syntax learning.

The rationale for a learning focus is this: although the capacity to use and understand words in novel positions undoubtedly requires generalization, and though the capacity to generalize *may* be largely innate, the *correct* use of words in novel positions necessitates *learning* to distinguish nouns from verbs and learning the argument structure of specific verbs. Thus, to achieve systematicity in a strong form, agents must integrate their generalization abilities with the syntax learning process. For this reason, both learning and generalization enter into my formulation of systematicity.

Now, systems vary in the degree to which they can generalize, and this variance will be central to the degrees of systematicity defined below. A question naturally arises as to how a system can be tested for various degrees of generalization in such a way that learning is involved. One approach, which is adopted here, is to distinguish between various degrees of *sentence novelty* that an agent can tolerate. For example, relative to sentences which comprise the agent's *learning base* (or training corpus), we can ask, what is the most novel sentence the agent can successfully interpret? If the agent has already encountered sentences which collectively contain some noun in every legal syntactic position, then a sentence that merely presents that noun in a novel combination with a known verb, does *not* constitute a strong test of the agent's capacity to generalize. By contrast, a sentence that presents some other, known noun in a novel syntactic position constitutes a stronger test.

2.2.1 Degrees of Systematicity

This section distinguishes three degrees of systematicity. No doubt, it is possible to make even finer distinctions, and in Section 5 I describe a still stronger, more human form of systematicity than any described in this section. However, for present purposes, the following should suffice. The three degrees of systematicity are: weak, quasi, and strong.

(1) *Weak Systematicity*. Networks exhibiting weak systematicity can perform at least the following kind of generalization. Suppose that a training corpus is 'representative' in the sense that every *word* (noun, verb, etc.) that occurs in some sentence of the corpus also occurs (at some point) in every permissible syntactic position. Thus, although the training corpus omits some sentences permitted by the target grammar, any network trained on this corpus will have been trained to recognize every word in every syntactic position that the word will occupy in the set of novel test sentences that are used to demonstrate the network's generalization capacity. Assuming that this set of novel sentences contains only grammatical sentences that are syntactically isomorphic to sentences in the training corpus, and that no new vocabulary is present, we shall say that a c-net exhibits at least *weak systematicity* if it is capable of successfully processing (by recognizing or interpreting) novel test sentences, once the c-net has been trained on a corpus of sentences that are *representative* in the sense described above. I describe such c-nets as (at least) weakly systematic in order to reflect the fact that their generalization capacity has only been tested upon sentences that are *weakly novel* with respect to the training corpus.

(2) *Quasi-Systematicity*. We shall say that a system exhibits quasi-systematicity if (a) the system *can* exhibit at least weak systematicity, (b) the system successfully processes novel sentences containing embedded sentences, such that both the larger containing sentence and the embedded sentence are (respectively) structurally isomorphic to various sentences in the training corpus, (c) for each successfully processed novel sentence containing a word in an embedded sentence (e.g. 'Bob knows that Mary saw *Tom*') there exists some *simple* sentence in the training corpus which contains that same word in the same syntactic position as it occurs within the embedded sentence (e.g. 'Jane saw *Tom*'). A system would be *merely* quasi-systematic if 'Tom' needed to occur (in the training corpus) in the *object position* of a *simple sentence,* before the system could correctly process embedded occurrences of 'Tom' in object position. Analogous remarks apply to subject position, verb position, etc.

(3) *Strong Systematicity*. We shall describe a system as strongly systematic if (i) it *can* exhibit weak systematicity, (ii) 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). Recall that a word can occupy the same syntactic position (e.g. subject) in both a simple and an embedded sentence. In such cases (where strong systematicity is concerned), neither the simple nor the embedded sentence counts as novel with respect to the other, unless other known words are being used in genuinely novel positions in those sentences. Also, given our background concern with cognitive plausibility, training corpora that 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. It is probably impossible to quantify, precisely, the relevant fraction, but it is well known that humans learn grammars on the basis of very incomplete data sets (cf. Pinker, 1989). (Not all the systems surveyed below can process embedded sentences. However, I shall argue on *independent grounds* that these systems exhibit only weak systematicity. None of the systems examined here attain strong systematicity, even when embedded sentences are excluded from consideration.)

By now it should be clear that strong systematicity, as just defined, shares much with F&P's conception of systematicity. Admittedly, there are important differences, but both conceptions seem to characterize roughly the same capacity to generalize. Having said that, I should add that the three degrees of systematicity I have introduced are primarily intended as technical aids to enable us to contrast specific abilities of existing c-nets with those of humans. I certainly do not wish to claim that my usage of 'weak systematicity', in particular, corresponds closely to any prior intuition the reader might have concerning the usage of that expression. Some readers might well feel that what I describe as 'weak systematicity' is too weak to be described as systematicity of any kind, since I do not require that a system that exhibits weak systematicity be sensitive to the selection restrictions (or case frames) of the verb, or that the system be context-sensitive in other ways (cf. Pinker, 1989). However, I invite the reader who feels this way simply to regard my 'weak systematicity' as a technical expression, for I certainly do not use the expression to suggest an ability comparable to human systematicity.

As it happens, c-nets are rather good at learning the selection restrictions of a verb, and are generally highly sensitive to context. Their shortcomings seem to lie in the opposite direction; they are so sensitive to context that they may be incapable of exhibiting strong systematicity, at least when cognitively plausible training regimes are employed. In what follows I hope to show that none of the c-nets surveyed here display anything stronger than quasi-systematicity. However, before examining these c-net results, I first argue that humans exhibit at least strong systematicity. As previously mentioned, an even stronger, more fully human form of systematicity is described in Section 5.

3. Humans and Strong Systematicity

To begin with, let us note that even young children, who have not yet reached the stage of producing multi-word utterances, are frequently able to obey *simple* imperative sentences which contain words in syntactic positions where the child has never encountered the word before. It is well known, for example, that in the few weeks which precede a child's first multi-word utterances, a 'spurt' occurs in a child's acquisition of *nominals* (both common and proper nouns), and that during this period children are able *rapidly* to acquire the use of nominals by means of 'what's that' games (cf. Ingram, 1989; Dromi, 1987). Once they have acquired nominals in this fashion, children are soon thereafter (i.e. within minutes) able to comprehend these words in sentences they encounter. This fact is established by Katz, Baker, and Macnamara (1974) who also present a strong case that the ability of young children to distinguish proper nouns from common nouns is much more a function of a child's *prior* ability to distinguish reidentifiable individuals from classes of objects than it is a function of some capacity to distinguish words that are syntactically preceded by an *article* from those which are not.

Additional evidence for strongly systematic behavior in children can be found in their use of passive formations, datavization, causative and locative constructions. For example, Pinker, Lebeaux, and Frost (1987) present several examples of spontaneous utterances by children in which the child coins a verb from a noun (as in 'It was bandaided'), and then uses the new verb in a passive formation. In many instances, it is implausible that the newly coined 'verb' has occurred anywhere in the child's training corpus, and *a fortiori* not in the passive position to which it is assigned by the child.

Pinker, Lebeaux, and Frost also conducted experiments in which children aged 3–7 years were presented with situations involving pairs of toys, and descriptions of these situations which employed novel (made-up) verbs (e.g., 'pilk' and 'gump'). All such descriptions employed the 'verbs' in active voice, and the children were able to surmise the verb-meanings via their ostensive uses (e.g. 'The tiger is pilking the horse'). Later, when the children were asked to describe new situations involving the toys they employed the 'verbs' in passive voice (since the focus of the situation was upon the patient of the action). Thus, the children were able to attach appropriate (passive) endings to the verbs, and to employ the passivized verb in a different syntactic context from that found in their training corpus.

Now, there is good reason to believe the children involved in these studies were capable of understanding their own utterances (as evidenced by their assent when the utterances were repeated by parents). Consequently, the *strong productivity* displayed by the children, both in the case of the newly-coined verbs and the nonsense syllable verbs, constitutes evidence of strong systematicity, as I have defined it. For in both cases the children are capable of understanding verbs when they occur in novel syntactic positions. Moreover, in the case of the child's coining a verb, the child is modifying a previously learned word and using the result in a position which is entirely novel, given the role of the original word.

Results analogous to those discussed above are presented in Gropen et al., 1989. Gropen et al. describe experiments involving the use of made-up verbs (such as 'pilk') in dative constructions (constructions involving both direct and indirect object). These experiments demonstrate children's ability to employ the novel verbs in double-object constructions (e.g. '*pilking* the pig the tiger') although the children have only heard the verbs used in dative contexts such as 'The bear is pilking the tiger to the pig'. Other work surveyed by Gropen et al. also reinforces the conclusion that children exhibit strong systematicity. Surveyed work includes many examples of spontaneous utterances in which children ungrammatically employ verbs in double-object constructions. For example, consider 'I go write you train' (from MacWhinney and Snow, 1985) and 'I'll brush him his hair' (from Mazurkewich and White, 1984). Now, as Pinker (1989) notes, 'Although some of these errors might have been caused by direct substitution of one verb stem for a semantically similar one (e.g. *write* for *draw* . . .) rather than by the application of a dative rule, most of them (e.g. *fix me a tiger*) must have involved the use of a rule'. Moreover, in cases such as 'I'll brush him his hair', we found nouns (e.g. 'hair') occurring in syntactic positions (second object) which they are unlikely to have occupied in any sentence the child has previously heard. The data surveyed by Gropen et al. contain several similar examples. Since the children involved understand their own utterances, we have good reason to believe children exhibit what I have called strong systematicity.

When we turn from children to adults, it becomes dramatically clear that humans can exhibit strong systematicity. Consider the following passage from Pinker, 1984:

> Furthermore, adults can be shown to be productive in their use of nouns heard in one position. Having heard the sentence *the romantizer was broken*, any adult can generalize to *I fixed the romantizer; the ROMANTIZER, I like; the romantizer's kadiddle; three romantizer-fixers; and so on.*

Note that even when we replace 'romantizer' with a word such as 'zorp', which has no prior semantic associations in English, Pinker's example still succeeds. Admittedly, Pinker's point concerns productivity. However, we could present any competent English user with just the sentence 'The zorp has returned', explain that a *zorp* is some living creature, and expect that person to understand both simple and embedded sentences containing 'zorp' in other syntactic positions. Since the person's grasp of 'zorp' is very limited, we would expect his/her understanding of the relevant sentences to be limited as well. Nevertheless, the person will be able to make reasonable sense of the sentences despite the fact that he/she associates little semantic content with 'zorp'.

4. The Connectionist Models

We turn now to consider connectionist systems which, prima facie, challenge Fodor and Pylyshyn's view on the limitations of c-nets *vis-à-vis* systematicity and structure-sensitive processing. As noted, not all these systems were *intended* as counterexamples to F&P's thesis, nor do their creators make strong claims for the psychological validity of their models. Nevertheless, in each case the model either performs generalizations related to systematicity or displays potential for structure-sensitive operations. It is revealing, therefore, to examine the ways in which each model measures up to the challenge of strong systematicity.

4.1 St. John and McClelland

St. John and McClelland (1990) present a connectionist model that learns to assign 'semantic representations' to English sentences which are presented as input. Although the details of their model are somewhat complex, the overall gist is that, via backpropagation, a network is trained to produce a correct semantic representation of the situation *described* by each input sentence. Situations (or events) described by input sentences consist of relationships, and the objects involved in those relationships.

Sentences that serve as input constitute a highly simplified version of English, in that all articles are deleted and only singular nouns are present. However, certain prepositional phrases are permitted. Input sentences are fed into the network in presegmented constituents. As each constituent is processed, an inspection is made, via a series of probes on an internal set of units, to see whether the network contains the desired, complete representation of the target situation. Backpropagation is performed after each such inspection. Because it is usually not possible to predict the *entire* target representation on the basis of isolated sentence constituents, the network is forced to learn associations between individual constituents and particular objects or relations in the target situation.

The network's input layer consists of a set of vocabulary units and a set of word-position units. Each sentence constitutent is represented in the input layer by activating one of the vocabulary units, corresponding to a noun, verb, etc., and a position unit. Vocabulary consists of nouns, verbs, prepositions, adverbs, and ambiguous words. Ambiguous words can function either as a noun or a verb, depending on context. The word-position units are used to indicate a constituent's position relative to the verb. Their inclusion may raise some eyebrows, but St. John and McClelland (hereafter, St.J&Mc) remark that they have obtained comparable results without using position units. Each complete sentence is presented to the network by sequentially activating pairs of units in the input layer, where each pair in the sequence corresponds to a word and word-position for a given sentence constituent.

The network's output, in response to a given constituent, consists of a set of values representing, respectively, thematic-role, concept (or word meaning), voice, and some subset of possible semantic features. There are nine thematic roles, including agent, action, patient, and location. As a well-trained network processes a sequence of constituents corresponding to an entire sentence, an appropriate corresponding sequence of representations for concepts, thematic-roles, and semantic feature values will appear on its output layer.

In addition to input and output layers, the network includes several internal layers of units. Two internal layers are of special interest. One of these, called the *sentence gestalt* (SG) layer, is eventually trained to produce a *distributed* meaning representation for each sentence in the training corpus. The other layer of interest, called the *context layer* retains a *copy* of the most recent contents of the SG layer. The SG contents change as each new sentence constitutent is fed into the input layer. During each iteration the contents of the input layer, together with the latest contents of the SG layer (which are retained in the context layer) are propagated upwards to the SG layer. The resulting new contents of the SG layer are then propagated upwards, but are also copied back to the context layer in readiness for the next iteration. The context layer enables the network to retain accumulated information about previous states of the network, and this accounts, in part, for the context-sensitive behavior displayed by the fully trained network.

Training the network involves, as each constituent is processed, repeatedly probing the sentence gestalt layer to see what series of values it produces on the output layer. Backpropagation is employed after each probe in response to errors revealed through probing. The exact manner in which probing and backpropagation are employed need not concern us, but it is noteworthy that the training procedure which is applied as each sentence constituent is propagated to the SG layer requires that the entire set of thematic-role/concept pairs, relevant to the meaning of the entire sentence, be known at the time backpropagation occurs. Thus, for the input sentence *'teacher drank tea'*, the set [(*agent/teacher*), (*action/drank*), (*patient/tea*)] must be known even at the time the first word of the sentence is processed.

In general, the training presupposes that the learning system possesses concepts applicable to all objects and relations in the target situation, and is also able to assign these concepts to differing thematic roles as the situation requires. Thus, in one situation the learner must recognize that 'teacher' assumes the agent role, and in another the patient role. What this means is that the learner must *already possess* a compositional, systematic method of conceptualization (representation) before the c-net training can even begin. Given this, St.J&Mc's model could not be used to explain how agents come to possess such an internal, compositional conceptual scheme. Moreover, nothing in St.J&Mc's overall design suggests that this conceptual scheme does not conform, in a general way, to the requirements of Fodor's (1976) Language of Thought thesis. Indeed, within the output layer of St.J&Mc's network, identical units are used to represent a given concept regardless of the role the concept plays on a given occasion. These concept units play an essential role in the training of the network. However, although St.J&Mc's model could not be used to depose Fodor's thesis (and this was certainly not their goal), the question still remains whether the model acquires knowledge of the compositional, systematic nature of its *input* sentences. St.J&Mc do not discuss the claims of F&P, but they clearly claim that such knowledge is acquired (1990, p. 250), and it is this claim we now consider.

As mentioned, St.J&Mc's training corpus includes sentences containing prepositional phrases. Unfortunately, when testing their network for the acquisition of compositional knowledge (which is manifested as *systematicity*) St.J&Mc used simpler training corpora, which lacked prepositional phrases. Two experiments were conducted to test for systematicity of behavior—one syntactic, the other semantic. Tests for syntactic systematicity involved only 10 objects and 10 reversible actions. Each object (action) uniquely corresponds to a particular noun (verb) in the training corpus. Both active and passive verb forms were permitted, and each input sentence had the general form: [noun verb-form noun]. Given that both active and passive forms are possible, a total of 2000 sentences are possible. All 2000 sentences were generated. Of these, 1750 comprised the training corpus, and the remaining 250 were set aside for later testing. Although St.J&Mc do not explicitly say so, their remarks elsewhere (1990, p. 243) suggest that these 250 sentences were randomly selected. Assuming they were, it is highly probable that the remaining 1750 sentences contained occurrences of every word in every legal syntactic position. (Otherwise, 80 per cent of the 250 test sentences would have to contain the *same* particular noun or verb in the same syntactic position. Given that there are 10 nouns and 10 verbs, this is extremely unlikely.[1]) Moreover, St.J&Mc give *no indication* that the training corpus (of 1750 sentences) *does not* include every possible word in every possible position. On the available

[1] Note that 10 per cent of the 2000 original sentences contain a given noun or verb in a given position. So, if a given noun or verb does not occur in a given position within the 1750 training sentences, then 200 of the 250 test sentences must contain that given word in the given position.

evidence, therefore, it is reasonable to believe that the training corpus does present every word in every possible position. This conclusion is reinforced by St.J&Mc's remark that 'What makes this a generalization task is that some of the sentences were set aside and not trained: *some agents were never paired with certain objects.*' (my emphasis) The fact that sentences in the *test* corpus describe novel agent–object combinations does present convincing evidence of generalization, but does not suggest that anything stronger than weak systematicity was tested for. To be sure, the network does display some degree of systematicity. The network assigns the correct semantic representation to 97 per cent of the novel 250 sentences. However, given the above considerations, it seems entirely likely that the network displays only *weak syntactic* systematicity.

The test for *semantic* generalization is analogous, in relevant respects, to the one just described. The semantic test involved a set of 400 possible sentences, of which 350 were used for training and the remaining 50 were used for testing. St.J&Mc explicitly note that the 50 test sentences were randomly selected from the set of 400. As before, the set of 400 sentences exhausts the space of possible sentences. Now, since the 50 test sentences were randomly chosen, it is extremely probable (by analogy with the reasoning given in the previous footnote) that each word occurred in a syntactic position within the *test* corpus that it also occupied within the training corpus. Thus, it is virtually certain that the test for *semantic* generalization established only weak systematicity. Certainly, we are given no reason to suppose otherwise.

It should be acknowledged, however, that despite the weaknesses mentioned above, the network we have considered yields some impressive results, including the ability to learn to 'disambiguate ambiguous words; instantiate vague words; assign thematic roles; and immediately adjust its interpretation as each constituent is processed.' (1990, p. 220) Even the ability to demonstrate weak systematicity is no small feat. However, it should be remembered that humans appear to exhibit a much stronger form of systematicity than this.

4.2 Elman

We turn now to the work of Elman (1989, 1990) on connectionist learning of syntactic structure. To begin, let us note that Elman clearly opposes his results (and those of others, including St.J&Mc) to the conclusions advanced by Fodor and Pylyshyn (1988), and to Fodor's (1976) Language of Thought thesis. In addition, Elman contends that 'the sensitivity to context which is characteristic of many connectionist models, and which is built-in to the architecture of the networks used here, does not preclude the ability to capture generalizations which are at a higher level of abstraction'. Yet, while it is clear that Elman's networks do generalize and acquire a degree of systematicity, it is doubtful that they display strong systematicity as defined here. Moreover, since Elman's research does not address issues of *semantic* systematicity and compositionality, it is unclear whether this work actually threatens Fodor's views on the Language of Thought. After all, we saw that St.J&Mc were able to train their network to discover *semantic* compositionality only when they assumed the prior existence of a compositional conceptual scheme. However, let us consider Elman's results in some detail.

Elman (1989) describes two experiments, both employing recurrent networks with a context layer feeding back into the hidden layer. The training procedure for both networks is essentially the same. Simplified English sentences (articles are absent) are fed into the network one word at a time, and backpropagation is used in a (prima facie) attempt to train the network to *predict* the next word it will receive as input. However, since a large training corpus is employed (10,000 sentences in each experiment), the network cannot learn to predict the next input word, but does learn (in essence) to predict the *syntactic category* of the following word. The first of the two experiments is designed, in fact, to demonstrate that the network does indeed develop a set of syntactic categories which correspond to the traditional grammatical categories. Cluster analysis on the network's hidden-layer activation values reveals that the network acquires *approximately* traditional categories, as well as approximate subcategories corresponding to animate noun, inanimate noun, transitive verb, etc.[2] The syntactic corpus for this experiment consists entirely of simple 2- and 3-word sentences. Both singular and plural nouns are included, and the network does learn to detect number agreement.

The second experiment is designed to test whether a somewhat more complex recurrent network can discover syntactic structure. In this experiment the training corpus includes relative clauses. Now, although the acquisition of *approximate* syntactic categories in the first experiment seems to indicate that a *degree* of systematicity has been discovered, only in this latter experiment is a test for systematicity explicitly performed. We therefore concentrate upon the latter experiment.[3]

The network in this case contains several internal layers, and a context layer. As before, the function of the context layer is to make the most recent state of one of the hidden layers available to that same hidden layer during the next input-output iteration. As a consequence, the network's context-sensitivity is enhanced.

Sentences are presented to the network one word at a time. The lexicon consists of 10 nouns, 12 verbs, the relative pronoun, 'who', and a *period*,

---

[2] These categories only approximate traditional categories because (for example) the representations developed for subject and object tokens of the same noun are not identical, though they do cluster together.

[3] Also, it is quite clear that the training corpus for the first experiment presented every word in every syntactically legal position. This can readily be established on the basis of the number of nouns and verbs available. It follows that the first experiment establishes only weak systematicity at best.

which is used to separate sentences in the input stream. Some of the 12 verbs optionally take a direct object. Unique bit vectors are used to encode each lexical item. Within the network, both input and output layers contain vectors of units equal in length to the lexical-item vectors. As each word in the input stream is presented to the input layer, and propagated upwards, some pattern appears on the output layer. The resulting pattern is compared to the encoding of the *next word* in the input stream and discrepancies determine the behavior of the backpropagation algorithm on the current iteration.

The training regime involves four phases, the first of which presents the network with a continuous stream of 10,000 sentences, containing no relative clauses. The three remaining phases each build upon the preceding phases, and involve increasingly high percentages of relative clauses. This controlled, graduated exposure to relative clauses naturally raises questions about psychological plausibility. However, in Elman, 1991, this gradual exposure is dropped in favor of a (seemingly) more plausible, gradual increase in the c-net's memory capacity. Under these altered conditions, Elman is able to obtain results comparable to those we are presently examining. More to our general point, though, in this recent work Elman does not address the issue of systematicity and does not discuss his training regimes in substantial detail. Certainly, no details are presented that would suggest that any stronger form of systematicity is achieved than in (Elman, 1989). Let us return, therefore, to the issue of systematicity in the 1989 paper.

Given that Elman's initial training phase involves 10,000 sentences, comprised of only 8 common nouns, 2 proper nouns, and 12 verbs, we have good reason to suppose that the *initial* training corpus presents every word in every syntactically legal position.[4] Assuming this is so, it is clear that the first phase could induce only weak systematicity. Moreover, there is no reason to suspect that any stronger form of systematicity is established by the later training phases, since every possible arrangement of nouns and verbs that could occur as the complement of a relative clause appears to be present within simple sentences in the first training corpus.

4.2.1 Integration of New Vocabulary

In Elman, 1990, the issue of integrating new vocabulary is briefly addressed. Elman describes an experiment in which his original training corpus (consisting of 10,000 simple (2- or 3-word) sentences) is modified as follows: all occurrences of 'man' are replaced by the *novel* word 'zog' throughout the corpus. This modified corpus is then presented to the original c-net, which had been trained on the original corpus. As the new

[4] Note that even if we assumed that every verb optionally takes a direct object, the total number of possible *simple* sentences is: [10 nouns × 12 verbs × 10 nouns = 1200] plus [10 nouns × 12 verbs = 120].

corpus is presented, no learning occurs, because the network's weights are frozen. However, as words are presented in sequence, hidden-layer activation patterns are collected and subjected to cluster analysis, just as in the original experiment. Now, although 'zog' has an *input* bit-encoding distinct form 'man', the cluster analysis reveals that the internal activation patterns assigned to 'zog', when averaged, bear the same relationship to other internal patterns, when averaged, as the patterns assigned to the word 'man'. In effect, 'zog' is assigned to the same subclass of nouns as 'man' previously occupied.

As Elman remarks, these results reveal that internal representations, assigned to various words, are very strongly affected by surrounding context. Given the strength of this contextual effect, it is natural to suppose that prior training could accelerate the network's acquisition of new vocabulary. We might also suppose that such *contextual effects* increase the likelihood that the network could correctly process *known words* in novel positions—thus creating the possibility of strong systematicity. However, this second supposition is hasty, as I now argue.

We should bear in mind that in the above experiment, 'zog' replaces 'man' in 10,000 sentences. 'Zog' is *never* presented to the network except where 'man' is expected, and the results of cluster analysis are based upon *averages* of activation values. These averages strongly reflect the contents of the *context layer*, which in turn reflects a sentential context identical to the context which accompanied 'man'.[5] Under these circumstances, it is hardly surprising that 'zog' is assimilated to 'man' by the c-net. By contrast, testing a c-net for strong systematicity can involve the presentation of very few novel sentences containing a known word that typically has appeared in positions uncorrelated with any other specific word. In this case, no significant opportunity for averaging of values exists, and the network is faced with words whose internal representations have already been developed. This is scarcely comparable to thousands of presentations of a word in positions *precisely identical* to some previously known word. Moreover, we must bear in mind that, in the 'zog' experiment, the c-net has already been trained upon a combinatorially exhaustive corpus. In particular, the word 'man' has been presented in every legal position during training. But, by definition, the training of a strongly systematic c-net cannot involve the presentation of all training-corpus words in every legal position.

4.3 Chalmers

Chalmers (1990) directly challenges Fodor and Pylyshyn's conclusions concerning the ability of c-nets to discover compositionality and exhibit

---

[5] I am indebted to Tony Plate for stressing the dramatic effect of the context layer on the averages used in cluster analysis. Indeed, Plate argues that the effects of the context layer are so strong that one can obtain a significant degree of clustering with a *completely untrained* network (personal communication).

systematicity. In Chalmer's view, Fodor and Pylyshyn seriously underestimate the potential for c-nets to perform structure mappings between *distributed* representations. He cites the work of Elman (discussed above) and Smolensky (1990, discussed below) as providing strong counterexamples to Fodor and Pylyshyn's claims. In addition, he describes his own experiment in which a network is trained to transform distributed representations of simple sentences in active voice into distributed representations in passive voice.

The vocabulary of Chalmer's training corpus consists of 5 proper nouns and 5 verbs. Active sentences have the form: N V N, while passives have the form: N is V by N. Initially, Chalmers generates all 250 active and passive sentences. These are divided into a *training* corpus of 150 sentences and a *test* corpus of 100 sentences. The training and test corpora are each equally divided between active and passive sentences.

Before any attempt is made to train the active–passive transformation into a network, Chalmers develops completely distributed representations for each sentence. Using techniques similar to Pollack (1990), Chalmers trains a 3-layer Recursive Auto-Associative Memory (RAAM) network to auto-associate on each sentence. Once trained, the RAAM will display a distributed pattern on its hidden (middle) layer in response to a given input sentence. Each such pattern is conventionally assigned as the representation of the sentence that produced it. A separate 3-layer network is then trained, using backpropagation, to perform the active–passive and passive–active transformations. Only distributed representations (developed during the RAAM training) are used as input and output in this second training phase.

The network involved in this latter phase is trained using 75 active- and 75 passive-voice sentence patterns, and within acceptable limits of error the network learns to output the appropriately transformed activation pattern. In a fashion similar to those examined above, Chalmers tests for acquisition of systematicity by means of the *novel* test corpus of 50 active and 50 passive sentences. However, all words occurring in this novel corpus occur in the initial training corpus, and within this initial corpus all words occur in all syntactically permissible positions (personal communication). Thus, in terms of the usage introduced in Section 2, it has only been shown that the network exhibits weak systematicity. While Chalmers establishes a convincing case that c-nets can perform mappings between distributed representations of structured objects, he has not demonstrated that c-nets can exhibit the kind of strong systematicity defined here.

In passing, it is worth noting that a variant of the 'fast-learning' phenomenon arises for the passive–active transformation. Suppose you are asked to give the passive form of 'Zim loves Zam'. Though you may pause momentarily to wonder about the unusual names, you are able to produce 'Zam is loved by Zim' without effort. Now, although context may provide a strong cue that 'Zim' and 'Zam' function as names, c-nets that have been trained to perform passive transformations in the fashion employed by Chalmers and St.J&Mc are certainly not able to produce *immediate* passive transformations upon distributed patterns they have never before encountered. In this connection, it is important that within both St.J&Mc's network, and Chalmer's, the internal distributed representation of 'loves' (after training) is context-sensitive, and that reading the novel word 'Zim' *may* put the network in a *novel* state in which it would not even be able to recognize 'loves' as a familiar lexical item. It is significant that humans are able to take an explicit instruction, such as 'Give me the passive form of X', and apply it to something which, for the agent, is at least a partially novel representation. I have argued elsewhere that this ability to apply transformations to internal representations *upon demand* (but not as an ordinary practice) suggests that a substantial degree of modularity is at work in behavior such as this.[6]

4.4 McClelland and Kawamoto

McClelland and Kawamoto (1986) present c-net results which, though predating the Fodor-Pylyshyn debate, might be seen as providing evidence of systematic behavior. Understandably, McClelland and Kawamoto (hereafter, Mc&K) make no claims about the compositionality or systematicity of their results. However, their c-net does generalize to some degree, and Mc&K maintain that their network is able to process sentences containing entirely novel words.

The overriding purpose of Mc&K's experiment is to train a network to learn associations between putative 'words' and the thematic case roles they assume in complete sentences. Their network contains two layers of units. On one layer a sequence of word representations may be presented, and on the other layer a set of thematic-roles may be represented. Training involves presenting the network with pairs of objects, having the form: (sentence-representation, corresponding thematic-roles), while a perceptron convergence algorithm is employed to ensure that each word within a given sentence representation eventually becomes associated with the thematic role played by that word in that sentence. As with St.J&Mc, the training process presupposes that the learner already possesses a compositional conceptual scheme containing concepts such as *agent*, *instrument*, etc.

An intriguing and indeed crucial aspect of the input representation is that sentences presented to the network contain not word-encodings *per se* but semantic encodings of words. Each word in an input sentence is replaced by a vector of semantic features representing, roughly, the word's sense. One way to think of this is that the 'spelling' of each word is the

[6] See Hadley, 1993, for more on the *rapid* application of explicit rules to internal representations.

word's corresponding vector of semantic features. Each verb has a vector of 6 semantic features; nouns have 8 features. Verb features include *doer*, and *causal*, while noun features include *human*, *form*, and *gender*. Each semantic feature is encoded by a unique bit vector, as are the thematic roles which are presented to the role layer. Thematic roles include *agent*, *verb*, *patient*, *modifier*, *instrument*, *food*, and *self*. Once the network has been fully trained, its performance can be tested by presenting a canonically represented sentence to the sentence layer, and observing that an appropriate sequence of role vectors is activated on the thematic-role layer.

Sentences that occur in the training corpus appear to be semantically representative of the *test* corpus in the following sense: although not every 'word' (semantic feature vector) is presented in every legal syntactic position during training, every *semantically relevant* subset of semantic features is presented in every legal position during training. Now, Mc&K maintain that once training is complete, their network is able to assign appropriate thematic roles to sentences containing *novel* words. Each 'novel word', however, is presented only in some sentence which is identical to a sentence in the training corpus, except that the novel word replaces a known word in that familiar sentence. Moreover, the semantic features that constitute the 'spelling' of the novel word are virtually identical to those of the replaced word. Thus, the putative novel word is, figuratively speaking, just a slight *misspelling* of the corresponding word in an otherwise familiar training sentence. In addition, the semantic features that have been altered in the novel word are not among those which are relevant to the thematic role in question. For example, in one test sentence the feature vector for the novel verb 'touch' is substituted for that of the previously acquired verb 'hit'. The two vectors differ only in a single feature, and this feature is irrelevant to the case frames which the network has learned.

Now, a problematic aspect of Mc&K's results is that all their successful examples of novel word recognition fit this pattern exactly. The 'novel words' are in reality feature vectors, which differ only slightly from the feature vectors of known words, and whose novel (discrepant) features are semantically irrelevant to the case frames in question. In light of this, I see no reason to suppose that Mc&K's results exhibit anything more than weak systematicity. Also, it is noteworthy that humans are sensibly able to employ novel words even when only a few semantic features are known. Even a three-year-old, upon being told only that 'a goblin is something very scary', would be able to form sentences such as, 'Daddy, will you keep away the goblins?' (Additional difficulties with Mc&K's method are described by Pinker, 1989, ch. 8.)

4.5 Pollack

To this point, we have examined four distinct systems, each of which is intended to induce either syntactic, semantic, or thematic knowledge from quasi-natural input. Thus far, we have found no reason to conclude that any of these systems exhibit strong systematicity. We turn now to the work of Pollack (1990) whose Recursive Auto-Associative Memory (RAAM) exhibits a degree of systematicity and productivity within a limited domain. Unlike the systems so far examined, Pollack's RAAM does not extract its knowledge from quasi-natural sentences, but from structured parse trees whose constituents are represented (in Lisp fashion) as nested lists.

As Pollack notes, the architecture of RAAM is essentially that of a simple auto-associative memory (Rumelhart et al., 1986). The network consists of equal-sized input and output layers, with a smaller hidden layer which develops condensed, distributed representations. Rumelhart et al. have demonstrated that simple networks of this kind can be trained, via back-propagation, to reproduce symbolic patterns presented to their input layers on their output layers. Merely reproducing patterns on an output layer is, of course, not very interesting. What is interesting is that such networks are able to replicate the input pattern on the basis of information contained in the intermediate, distributed representation. Pollack takes the work of Rumelhart et al. a few steps further by demonstrating that networks of the above kind can be trained to replicate *nested* structures, such as parse trees, on their output layers. For example, consider the following parse tree which is represented by means of nested lists:

((Det (Adj Noun)) (Verb (Prep (Det Noun))))

A RAAM network can be recursively trained, in stages, to auto-associate a bit-form representation of this tree structure with itself. The training involves *first* training the network to auto-associate on the simplest constituents, and then on progressively more complex constituents. In the final stage, the network is trained to auto-associate on the entire parse tree. Thus, for the example above, the network is trained to auto-associate on the simple constituents, (Adj Noun) and (Det Noun), before being trained on (Det (Adj Noun)), (Prep (Det Noun)), (Verb (Prep (Det Noun))) and ultimately on the entire expression.

In one experiment, which demonstrates the productivity and systematicity of RAAM, the network is trained to auto-associate on 13 *ternary tree* structures, which encode 13 English sentences of increasing complexity. (Generally speaking, ternary trees may be represented as triads of the form, (α β γ), where α, β and γ may each be an atomic symbol or another triad of similar form. However, Pollack's application requires that α will always be atomic (e.g. a verb), since Pollack encodes relational terms in first position.) An example of such a ternary tree, which includes an embedded structure, is:

(THOUGHT PAT (KNEW JOHN (LOVED MARY JOHN))).

The 13 tree structures, which comprise the input set, range in complexity from

(LOVE PAT MARY)

to

(SAW (IS (MOD MAN SHORT) (THOUGHT MAN (SAW MAN JOHN))) PAT).

Each terminal symbol is encoded by a 16 bit vector, and bit encodings of parse trees are built up out of these vectors. Backpropagation is employed to ensure that the 3-layer RAAM is capable of reproducing its input pattern upon its output layer, within a reasonable margin of error.

Once the network is trained on the 13 ternary trees, it is tested for its ability to auto-associate on *novel* trees, which correspond to *novel* sentences comprised of the same atomic constitutents. The RAAM network succeeds in this task, though syntactically mal-formed trees are occasionally generated. Certain of the *successful* trials present noun–verb combinations which are novel in the sense that the particular noun does not appear as an argument to that verb anywhere in the training corpus. Thus, the network has unquestionably induced some of the compositional structure implicit in the training corpus. However, Pollack's sole argument for systematicity rests upon the fact that 'All 16 cases of (LOVED X Y), with X and Y chosen from the set JOHN, MARY, PAT, MAN were able to be reliably represented, even though only four of them were in the training set'. As it happens, although one of the words, 'man', never appears as argument to 'loves' in the training corpus, each of the four nouns occurs as subject and as direct object somewhere in the training corpus. Thus, *strong* systematicity appears not to have been established. A case for a kind of quasi-systematicity can be made, however, since within the training corpus 'John' occurs as a direct object only within an embedded constituent. Yet, the network successfully processes the novel sentence, (LOVED MARY JOHN), which corresponds to the simple English sentence, 'Mary loved John'. (Quasi-systematicity also occurs in the recent results of Niklasson and Sharkey (1992), whose training methods combine those of Chalmers and Pollack.[7])

[7] The strongest example of systematicity described by Niklasson and Sharkey involves explicitly parenthesized logic formulae, comparable with Pollack's nested lists. Their network is able to perform logic transformations on distributed encodings of formulae such as (atomic-wff → (p connective atomic-wff)), although the network had never been trained to transform formulae that contain the constant p in precisely the syntactic position it occupies in the schema displayed above. Nevertheless, though the network was never trained for this specific task, it was trained on formulae isomorphic to this schema, as well as on *simple* sentences of the form (p connective atomic-wff) (personal communication). It follows that Niklasson's and Sharkey's experiment exhibits what I have called quasi-systematicity.

Turning to considerations of cognitive validity (which, admittedly, was not one of Pollack's primary concerns), a difficulty with Pollack's method arises, viz., that the RAAM network is trained *only* on pre-parsed, explicitly parenthesized structures whose syntactic structure has *already* been analyzed. Thus, his training regime, like that of St.J&Mc, is parasitic upon an external agent's having already discovered the systematic properties of the input set. Pollack displays awareness of this difficulty when he notes, 'The simplifying assumption, that internal representations can first be devised and then used as target patterns, is questionable'. Given that Pollack's primary concern was not cognitive fidelity, but to demonstrate RAAM's capacity for structure-sensitive processing, his simplifying assumption is understandable. However, within the context of the long-term goals of cognitive science, this simplification is indeed problematic.

4.6 Smolensky

In Smolensky (1990) the question whether highly structured, compositional representations can be realized in completely distributed form is examined from a mathematical perspective. Smolensky's answer is affirmative; he is able to demonstrate with considerable rigour, using tensor calculus as a tool, that concatenative symbolic representations (including representations of nested structures) can be translated into tensor product representations. These representations subsume the vector space representations commonly employed in the analysis of c-nets, and Smolensky demonstrates that *many* existing connectionist systems can straightforwardly be analyzed using the tensor calculus.

Smolensky's general strategy is to show (1) that structured symbolic representations can be analyzed as *sets* of filler/role pairs, (2) that each such set has a tensor product representation, and (3) there exist systematic ways of instantiating tensor product representations in c-nets, in a *distributed* fashion. The idea underlying point (1) is simple but effective. Given an n symbol string, each symbol, s_i, fills the role of occupying position r_i in the string. Thus, the string '*c f c*' is represented in filler/role format by the *unordered set* $\{ c/r_1, f/r_2, c/r_3 \}$. Of the three points in Smolensky's strategy, perhaps (3) bears most directly upon the Fodor–Pylyshyn thesis. For, since the tensor product representations correspond to conventional, structured representations, Smolensky contends that Fodor and Pylyshyn grossly underestimate the power of *distributed* connectionist representations.

Fodor and McLaughlin (F&McL) have challenged Smolensky's stand on this issue, but their argument, in my view, is problematic. For F&McL *assume* it to be nomologically necessary that cognitive agents exhibit systematicity. (Something is nomologically necessary if it is necessary by virtue of natural law. For convenience, I will use 'nomically-arbitrary' to denote anything which lacks necessity of this kind.) They concede that, using tensor product representations, Smolensky could *hard-wire* a c-net to exhibit systematicity, but they regard such wiring as purely arbitrary (cf. F&McL, 1990, p. 202). They emphasize that Smolensky could just as well hard-wire a c-net to produce output that is *inappropriately* related to input as anything else. F&McL do not say whether they regard innate wiring, which arises through natural selection, as nomically-arbitrary, but the structure of their argument seems to require that they do so. Otherwise, Smolensky might plausibly rejoin that his systematicity-inducing hard-wiring corresponds, for all we know, to some innate (non-arbitrary) mechanism. Given this, it seems safe to conclude that F&McL are precluded from regarding the existence of innate mechanisms as nomologically necessary. However, this *preclusion* undermines F&McL's case on a different front (as I now argue).

F&McL repeatedly assert, in essence, that even assuming classical constituent structure, an agent will display systematicity only if *specific* kinds of structure processing mechanisms are in place. Such mechanisms must either be innate, or the product of training, or some mixture of the two. Given that F&McL seem committed to describing innate mechanisms as nomically-arbitrary, and given that training is a contingent event, it is difficult to see how F&McL could consistently regard the existence of structure-processing mechanisms as nomologically necessary. Moreover, if the existence of these mechanisms is not necessary in this sense, then why should we grant F&McL that systematicity is the product of nomic necessity?

In any case, given the conception of systematicity I have proposed, it is clear that agents exhibit systematicity (in my sense) only after a period of learning. Thus, even if one accepts Fodor's and McLaughlin's arguments (using their sense), I could not avail myself of their arguments. For this reason, my discussion of Smolensky's work proceeds along somewhat different lines from Fodor's and McLaughlin's, though there are points of overlap.

As previously noted, both Chalmers and Smolensky maintain that Fodor and Pylyshyn have underestimated the power of distributed representations. Whether or not one accepts this claim, it is far from clear that Smolensky's methods establish that c-nets can exhibit strong systematicity. To see this, let us consider how a c-net might come to possess distributed representations of the kind Smolensky describes. For the most part, Smolensky sidesteps this issue, as when he says (1990, p. 193): 'Here it is not the job of the network to set up the tensor product representations: in presenting the input/output pairs to the network during training, the modeler must convert the symbolic inputs and outputs to their vector representations, and this can be done directly by using the mathematical definition of the tensor product representation'. Elsewhere, Smolensky explains how to construct the required vector representations from tensor product representations, but what we are told, in effect, is how to hand-construct a series of highly regular, grid-like structures whose dimensions are determined by the number of fillers and roles involved. For example, if a symbolic expression has been represented using three fillers and three roles, then we may construct a grid-like network of the kind shown below.

Smolensky provides alternative methods for creating similar grids, but the kind shown below is as simple as any of the alternatives. In the network shown, each filler unit sends out horizontal connections and each role unit sends out vertical connections. A binding unit resides at each point of intersection, *(i,i)*, and makes connections with f_i and r_i. If we assume that in a given representation f_3 is to fill role r_3, then activation at binding node *(3,3)* represents the binding of the filler to the role, f_3/r_3.

Smolensky offers no suggestion as to how a particular set of binding activations, which represent an entire sentence, might arise through a learning process. Near the end of his paper, Smolensky briefly discusses how sets of *roles* (which occur in the sets of filler/role pairs comprising tensor product representations) might be learned. In particular, he mentions that distributed representations corresponding to individual roles may be developed by means of the backpropagation algorithm. In certain cases (where representations for filler and role vectors are not *linearly independent*), error can arise in the decomposition of complex representations into constituent parts. For such cases, Smolensky has demonstrated how gradient descent techniques can be used to develop role representations which minimize error in the decomposition process. However, there is no indication that any of these learning processes are more powerful than those employed by Elman (whose networks also learn grammatical *roles*), and we have already seen that Elman's method appears to establish only weak systematicity. While it is imaginable that Smolensky's distributed representations of filler/role pairs could be acquired without the

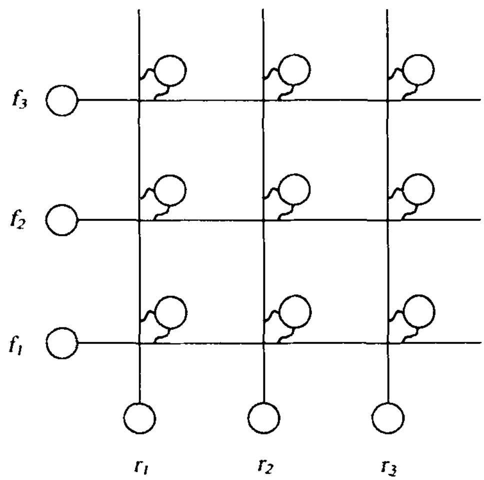

Figure 1

kinds of exhaustive training regimes we have examined thus far, we have at present no evidence which would decide this issue.

Of course, it might be suggested that humans, at least, innately possess elements that function as role nodes and filler nodes, and innately possess the ability to produce appropriate sets of activation bindings. However, even assuming the innateness of all this, a serious difficulty yet remains. For, whether or not humans possess these things innately, we should bear in mind that static representations cannot, by themselves, create systematicity. A system exhibits systematicity only if it can generalize in specific ways, and this entails the ability to *process* its representations. C-nets can process their representations only if the appropriate weight vectors are in place, and Smolensky presents no arguments to show that these weight vectors can be acquired in a manner that avoids the exhaustive training regimes we have already examined. This is understandable. Smolensky was not addressing a learning-based conception of systematicity, much less strong systematicity as I have defined it. Indeed, Smolensky's recent work (Miyata, Smolensky and Legendre, 1993) on structure-sensitive processing employs tensor representations in the context of *hard-wired* networks. Unfortunately, these hard-wired implementations shed little light on the question whether c-nets can exhibit strong systematicity, as conceived here.

Apart from this, given the nature of Smolensky's tensor product representations (regularity, predictability, and isomorphism[8] with traditional symbolic structures) questions arise whether hard-wired manipulations of these structures should not be viewed as *implementations* of classical symbolic processing (cf. Fodor and McLaughlin, 1990).

5. Discussion

In the foregoing I have examined a number of connectionist experiments, due to St. John and McClelland, Elman, Chalmers, McClelland and Kawamoto, Pollack, and Smolensky. With the possible exception of Mc&K's results, each of these experiments has been depicted, by one or more of the foregoing, as undermining the conclusions advanced by F&P *vis-à-vis* compositionality and systematicity. Whether these conclusions are indeed undermined, I leave to others to determine. However, I suspect that F&P would have sympathy for the conception of strong systematicity offered here, and it appears clear that none of the systems examined here displays

[8] Note that Smolensky's method of representation maps linear symbolic sequences into a higher dimensional space. The mapping is straightforward, and whenever decomposition can be reliably performed, the mapping is reversible. Thus, when Smolensky's method produces *adequate* representations (i.e. decomposition is feasible) the syntax of such representations can be placed in a direct isomorphism with that of conventional symbolic strings.

strong systematicity as presently defined. In the work of St.J&Mc, Chalmers, and Mc&K we found no evidence that anything stronger than weak systematicity had been achieved. Smolensky's results, which clearly establish the capacity for structure-sensitive operations in c-nets, do not involve learning methods, and thus cannot, presently, be placed upon the spectrum of systematicity given in section two. In Pollack's work (as well as Niklasson and Sharkey's, see n. 7) we did find evidence of quasi-systematicity, but even this falls short of strong systematicity. Moreover, Pollack's method requires that structured, spatially concatenated representations be presented to the networks at the time that backpropagation is employed. His method thus assumes the existence of an external agent who has already discovered and encoded the compositional structure of a set of representations (as does St.J&Mc's and Mc&K's).

We should bear in mind, also, that humans may well exhibit stronger forms of systematicity than any defined thus far. For example, humans have the capacity not only to recognize words in novel syntactic positions, but also to assign appropriate meanings to those words and the resulting novel sentences. Because several of the experiments examined here address only syntacatic generalization (and Elman's network in particular generalizes only in its capacity to predict the grammatical *category* of succeeding words of input), I have chosen not to include semantic aspects in my definition of strong systematicity. Yet, I think it clear that human systematicity does include semantic aspects, and that a c-net might satisfy my definition of strong systematicity and still fail to distinguish the *meanings* of 'John loves Mary' and 'Mary loves John'. By requiring that agents be able to assign appropriate meanings to relevant novel sentences, in addition to satisfying strong systematicity, we arrive at a fourth (yet more human) degree of systematicity.

Having said all this, I do not wish to suggest the results considered here are uninteresting. On the contrary, I suspect that they will furnish important clues to the riddle of language acquisition. (Also, I have argued elsewhere (Hadley, 1989) that c-nets may provide a key component of *semantic grounding* processes.) I submit, however, that c-nets will exhibit strong *semantic* systematicity only when systems of c-nets are devised, which can exploit the fact that tokens of the same word, appearing in different syntactic positions, can have the same *context-free* semantic content. Such systematicity may very well require that tokens of the same word which occur in distinct syntactic locations be *recognized* as belonging to the same type even when the given token has never appeared in that syntactic position before. This, in turn, seems to require that context-free *token recognizers* be embedded in c-net systems. However, if the necessity for context-free recognizers is granted, and if we grant (as it seems we must) that language learners must *somehow* respond to structural differences in sentences that contain the same word in differing positions, it would seem that one of Fodor and Pylyshyn's conclusions (that human compositionality and systematicity entail the existence of structure-sensitive operations on context-free objects) stands unscathed. Of course, this last conclusion is conditionalized, and I have offered no proof of its antecedent premises.

What I *have* argued, though, is that none of the experiments we have examined display the degree of systematicity characteristic of humans. The highest degree of c-net systematicity we have seen is quasi-systematicity, and this has only been achieved when backpropagation of error has been rigorously applied and when distributed representations were developed for pre-parsed, fully parenthesized sentences. The fact that quasi-systematicity has only been demonstrated under conditions of questionable cognitive validity does not suggest that *strong* systematicity can be achieved under cognitively realistic conditions, at least when the kinds of non-modular, connectionist methods examined here are the means employed.

School of Computing Science and Cognitive Science Program
Simon Fraser University
Burnaby
B.C., V5A 1S6
Canada

References

Berwick, R.C. 1986: *The Acquisition of Syntactic Knowledge*. Cambridge, MA.: MIT Press.

Chalmers, D. 1990: Why Fodor and Pylyshyn Were Wrong: the Simplest Refutation. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, Cambridge, MA.

Dromi, E. 1987: *Early Lexical Development*. Cambridge University Press.

Elman, J.L. 1989: Representation and Structure in Connectionist Models. CRL Technical Report 8903, University of California at San Diego, La Jolla, CA. 92093-0126.

Elman, J.L. 1990: Finding Structure in Time. *Cognitive Science*, 14, 179–212.

Elman, J.L. 1991: Incremental Learning, or the Importance of Starting Small. CRL Technical Report 9101, Center for Research in Language, University of California at San Diego, La Jolla, CA. 92093-0126.

Fodor, J.A. 1976: *The Language of Thought*. Cambridge, MA.: Harvard University Press.

Fodor, J.A. and McLaughlin, B.P. 1990: Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work. *Cognition*, 35, 183–204.

Fodor, J.A. and Pylyshyn, Z.W. 1988: Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 28, 3–71.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R. and Wilson, R. 1989: The Learnability and Acquisition of the Dative Alteration in English. *Language*, 65, 203–57.

Hadley, R.F. 1993: Connectionism, Explicit Rules, and Symbolic Manipulation. *Minds and Machines*, 3, 183–200.

Hadley, R.F. 1989: A Default-Based Theory of Procedural Semantics. *Cognitive Science*, 13, 107–37.

Ingram, D. 1989: *First Language Acquisition*. Cambridge University Press.

Katz, N., Baker, E. and Macnamara, J. 1974: What's in a Name? A Study of How Children Learn Common and Proper Names. *Child Development*, 45, 469–73.

Mazurkewich, I. and White, L. 1984: The Acquisition of the Dative Alteration: Unlearning Overgeneralizations. *Cognition*, 16, 261–83.

McClelland, J.L. and Kawamoto, A.H. 1986: Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group (eds), *Parallel Distributed Processing; Explorations in the Microstructure of Cognition, Volume 2*. Cambridge, MA.: MIT Press.

MacWhinney, B. and Snow, C. 1985: The Child Language Data Exchange System. *Journal of Child Language*, 12, 271–96.

Niklasson, L. and Sharkey, N.E. 1992: Systematicity and Generalization in Connectionist Compositional Representations. Technical Report RR-92-01-005, Department of Computer Science, University of Skövde, Sweden.

Pinker, S. 1984: *Language Learnability and Language Development*. Cambridge, MA.: Harvard University Press.

Pinker, S. 1989: *Learnability and Cognition: the Acquisition of Argument Structure*. Cambridge, MA.: MIT Press.

Pinker, S., Lebeaux, D.S. and Frost, L.S. 1987: Productivity and Constraints in the Acquisition of the Passive. *Cognition*, 26, 195–267.

Pollack, J.B. 1990: Recursive Distributed Representations. *Artificial Intelligence*, 46, 77–105.

Rumelhart, D.E., Hinton, G. and Williams, R. 1986: Learning Internal Representations through Error Propagation. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1*. Cambridge, MA.: MIT Press.

Smolensky, P. 1990: Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems. *Artificial Intelligence*, 46, 159–216.

Miyata, Y., Smolensky, P. and Legendre, G. 1993: Distributed Representation and Parallel Processing of Recursive Structures. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, Boulder, Colorado, 759–64.

St. John, M.F. and McClelland, J.L. 1990: Learning and Applying Contextual Constraints in Sentence Comprehension. *Artificial Intelligence*, 46, 217–257.

Van Gelder, T. 1990: Compositionality: A Connectionist Variation on a Classical Theme. *Cognitive Science*, 14, 355–84.