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.

Article
Generalization and Connectionist Language Learning

MORTEN H. CHRISTIANSEN AND NICK CHATER

The performance of any learning system may be assessed by its ability to generalize from past experience to novel stimuli. Hadley (this issue) points out that in much connectionist research, this ability has not been viewed in a sophisticated way. Typically, the 'test-set' consists of items which do not occur in the training set; but no attention is paid to the degree of novelty of test items relative to training items. Hadley's arguments challenge connectionists to go beyond using training and test sets which are chosen according to convenience, and to carefully construct materials which allow the degree of generalization to be investigated in more detail. We hope that this challenge will encourage more sophisticated connectionist approaches to generalization and language learning.

Hadley defines different degrees to which a language learning system can generalize from experience, what he calls different degrees of systematicity. In this paper we discuss and attempt to build on Hadley's account, providing more formal and precise definitions of these varieties of generalization. These generalizations aim to capture the cases that Hadley discusses, but also extend to other examples. We then report some connectionist simulations using simple recurrent neural networks which we assess in the light of the revised definitions. Finally, we discuss the

Address for correspondence: Morten H. Christiansen, Philosophy-Neuroscience-Psychology Program, Department of Philosophy, Washington University, One Brookings Drive, Campus Box 1073, St. Louis, MO 63130-4899, USA. Nick Chater, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK. Email: morten@twinearth.wustl.edu; nicholas@cogsci.ed.ac.uk.

prospects for connectionist and other approaches to language learning for meeting these criteria, and consider implications for future research.

1. Systematicity and Generalization

1.1 Varieties of Systematicity

Hadley defines several levels of systematicity, which are increasingly difficult for a learning system to meet. Following Hadley and emphasizing the learning of *syntactic* structure, we focus on the first three, weak, quasi- and strong systematicity, as benchmarks for current connectionist models ('c-net' in Hadley's terminology).

According to Hadley, '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 which are *representative*' p. 250). A training corpus is 'representative' if 'every *word* (noun, verb, etc.) that occurs in some sentence of the corpus also occurs (at some point) in every permissible syntactic position' (p. 250). *Quasi-systematicity* can be ascribed to a system 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*")' (p. 250). Finally, a system will exhibit *strong systematicity* 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)' (p. 250).

Central to each definition is the notion of *syntactic position*, which may or may not be shared between items in the training and test sets. Since syntactic position is not a standard term in linguistics, and since it is not discussed in the paper, we must examine Hadley's examples to discover what meaning is intended. These are concerned with the relationship between verbs and their arguments. The various argument positions of a verb (subject, direct object and indirect object) are taken to count as distinct syntactic positions. Also, the active and passive forms of a verb are taken to occupy different syntactic positions.

If these examples are taken at face value, difficulties emerge. For example, a lexical item is the subject with respect to some verb whether or not it

occurs within an embedded sentence, a simple sentence, or the main clause of a sentence which contains an embedded sentence (and similarly with the other examples). This means that, for Hadley, '*John*' has the same syntactic position in '*John loves Mary*' as in '*Bill thinks that John loves Mary*'—indeed, this is explicit in point (c) of the definition of quasi-systematicity. Nonetheless, it would appear that, according to Hadley, a learning system which generalizes from either of these sentences to the other only requires weak systematicity (since no item occurs in a novel syntactic position). Yet, this seems to be exactly the kind of case which is supposed to distinguish quasi-systematicity from weak systematicity in Hadley's definitions. But, as we see, it appears that weak systematicity already deals with such cases, if syntactic position is defined in terms of grammatical role, since grammatical role abstracts away from embedding. Quasi- and weak systematicity therefore appear to be equivalent.

Presumably, either weak or quasi-systematicity is intended to have an additional condition, which is not explicit in Hadley's definition. We suggest one possible condition below, that quasi-systematicity is only exhibited when the test and training sets contain embedded sentences. An alternative interpretation would be that Hadley is implicitly making use of a more global notion of syntactic context, which distinguishes the syntactic position of a subject in a sentence which contains an embedded clause, and one that does not, for example.[1]

In order to extend the account beyond the cases of subject and object, we need some more general account of syntactic position. We suggest a possible definition below, and use it to define what we call three levels of generalization, which we intend to be close to the spirit of the original definitions of systematicity.

1.2 Syntactic Context

The syntactic position of a word is defined in terms of the phrase structure tree assigned to the sentence in which it occurs. We use phrase structure trees since they are linguistically standard and can be used in a precise and general way. We intend no theoretical commitment to phrase structure based approaches to linguistic theory. Our account could be given equally well in alternative linguistic frameworks.

We define the syntactic position of a word to be the tree subtended by

---

[1]   Hadley (personal communication) seems to lean towards the latter interpretation in a recent version of his definition of weak systematicity: 'the training corpus used to establish weak systematicity must present every word in every syntactic position and must do so at all levels of embedding found in the training and test corpus. In contrast, a quasi-systematic system does not have to meet the condition in the second conjunct, but does satisfy the first conjunct'. Notice that this revision suggests that Elman's (1989, 1991a) net might be quasi-systematic after all (pace Hadley, this issue, p. 259).
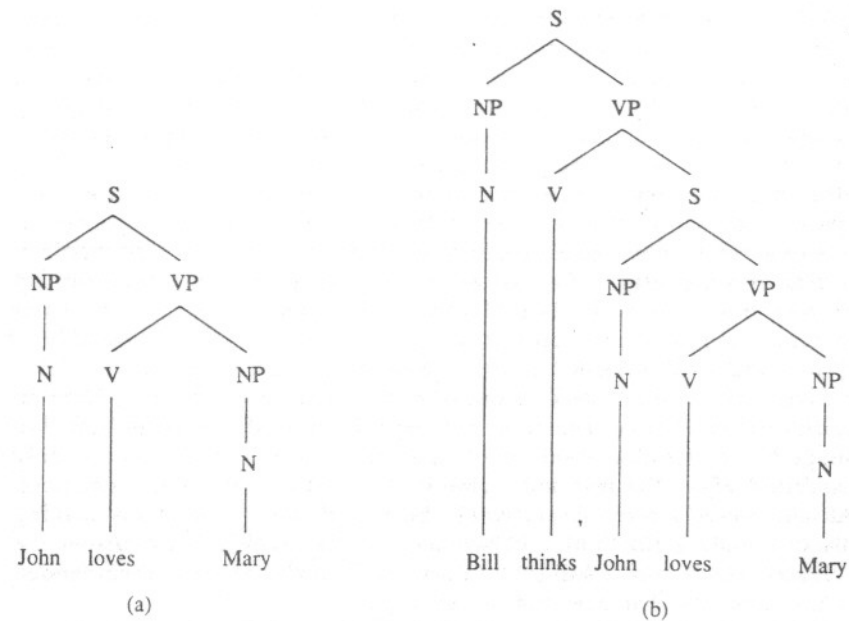
Figure 1 *Phrase structure trees for (a) the simple sentence* 'John loves Mary' *and (b) the complex sentence* 'Bill thinks John loves Mary'.

Figure 2 *The syntactic position of (a) the subject noun and (b) the object noun in the sentence* 'John loves Mary'.

Figure 3 *The syntactic position of (a) the main verb and (b) the subordinate verb in the sentence* 'Bill thinks John loves Mary'.

the immediately dominating S or VP node, annotated by the position of the target word within that tree. This tree will be bounded below either by terminal nodes (Det, Proper Noun, etc), or another S or VP-node (i.e. we do not expand the syntactic structure of embedded sentences or verb phrases).

For example, consider the phrase structure trees for the simple sentence 'John loves Mary' and the complex sentence '*Bill thinks John loves Mary*' as shown in Figure 1. In a simple sentence like 1 (a), the subject is defined by its relation to the dominating S-node. The object and the verb are defined in relation to the verb phrase. This captures the distinction between subject and object noun positions. Figure 2 (a) and (b) depict this distinction, illustrating, respectively, the syntactic positions of '*John*' and '*Mary*'.

Also according to this definition, verbs with different argument structure are considered to have different syntactic contexts. For example, intransitive, transitive and ditransitive occurrences of verbs will be viewed as inhabiting different contexts. Furthermore, verb argument structure is relevant to the syntactic context of the object(s) of that verb, but not of its subject.
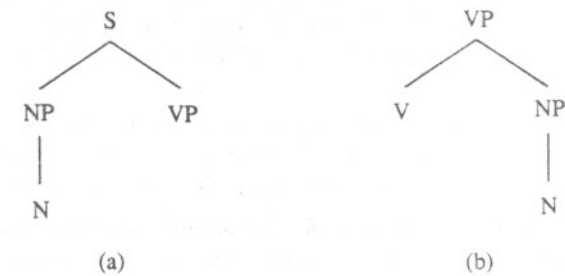
In a complex sentence like 1 (b), there will be different local trees for items in the main clause or in any embedded clauses. For example, '*thinks*', which occurs in the main clause of 1 (b), has a syntactic position defined with respect to the verb phrase pictured in Figure 3 (a), whereas for '*loves*' in the embedded clause, the syntactic position is defined with respect to the structure of the embedded sentence shown in 3 (b). The two trees in Figure 3 are thus examples of how the verb argument structure affects syntactic position.

Notice that this means that the syntactic position within an embedded clause is affected only by its local context, and not by the rest of the sentence. Thus the notion of syntactic position applies independently of the depth of embedding at which a sentence is located. Furthermore, according to this definition, the syntactic context of a word in a particular clause is not affected by the structure of a subordinate clause; and the syntactic context of a word in a subordinate clause is not affected by the structure of the main clause.

1.3 Varieties of Generalization

Using this definition of syntactic position, we can now recast Hadley's definitions to give three levels of generalization for language learning systems.

1. *Weak Generalization*: A learning mechanism weakly generalizes if it can generalize to novel sentences in which no word occurs in a novel syntactic position (i.e. a syntactic position in which it does not occur during training).[2]

2. *Quasi-Generalization*: A learning mechanism is capable of quasi-generalization if it can generalize to novel sentences as in (1), with the additional constraint that embedding occurs in the grammar.

3. *Strong Generalization*: A learning mechanism strongly generalizes if it can generalize to novel sentences, that is, to sentences in which some (sufficiently many) words occur in novel syntactic positions.

This definition of strong generalization, implies that for the two test sentences:

> *John thinks Bill loves Mary.*
> *Bill loves Mary.*

if '*Mary*' had never occurred in the object position in the training set (in either embedded or main clauses), the syntactic position of '*Mary*' in both these sentences would be novel. If '*Mary*' had occurred in object position at all in the training set, then in neither sentence is the syntactic position novel.

These definitions aim to capture the spirit of Hadley's proposals in a reasonably precise and general way. We now turn to some simulations which aim to test how readily these definitions can be met by a simple recurrent network.

2. Simulations

As a first step towards meeting the strong generalization criterion described above, we present results from simulations involving a simple recurrent network. The research presented here (and elsewhere, e.g. Chater, 1989; Chater and Conkey, 1992; Christiansen, 1992, in preparation; Christiansen and Chater, in preparation) builds on and extends Elman's (1988, 1989, 1990, 1991a, 1991b) work on training simple recurrent networks to learn grammatical structure. Hadley (this issue, p. 259) rightly notes that the training regime adopted by Elman (1988, 1989, 1990) does not afford strong systematicity (nor does it support our notion of strong generalization) since the net by the end of training will have seen all words in all possible syntactic positions. We therefore designed a series

---

[2] Note that Hadley's revised definition of weak systematicity (as mentioned in the previous footnote) differs from this notion of weak generalization.

```
S     → NP VP "."
NP    → PropN | N | N rel | N PP | gen N | N and NP
VP    → V (NP) | V that S
rel   → who NP VP | who VP
PP    → prep NP
gen   → N + "s" | gen N + "s"
```

Figure 4 The phrase structure grammar used in the simulations

of simulations aimed at testing how well these nets can capture strong generalization.

In our simulations, we trained a simple recurrent network to derive grammatical categories given sentences generated by the grammar shown in Figure 4. This grammar is significantly more complex than the one used by Elman (1988, 1991a). The latter involved subject noun/verb number agreement, verbs which differed with respect to their argument structure (transitive, intransitive, and optionally transitive verbs), and relative clauses (allowing for multiple embeddings with complex agreement structures). We have extended this grammar by adding prepositional modifications of noun phrases (e.g. '*boy from town*'), left recursive genitives (e.g. '*Mary's boy's cats*'), conjunction of noun phrases (e.g. '*John and Mary*'), and sentential complements (e.g. '*John says that Mary runs*'). Our training sets consisted of 10,000 sentences generated using this grammar and a small vocabulary containing two proper nouns, three singular nouns, five plural nouns, eight verbs in both plural and singular form, a singular and a plural genitive marker, three prepositions, and three ('locative') nouns to be used with the prepositions. A few sample sentences are in order:

> *girl who men chase loves cats.*
> *Mary knows that John's boys' cats eat mice.*
> *boy loves girl from city near lake.*
> *man who girls in town love thinks that Mary jumps.*
> *John says that cats and mice run.*
> *Mary who loves John thinks that men say that girls chase boys.*

To address the issue of generalization, we imposed an extra constraint on two of the nouns (in both their singular and plural form). Thus, we ensured that '*girl*' and '*girls*' never occurred in a genitive context (e.g. neither '*girl's cats*' nor '*Mary's girls*' were allowed in the training set), and that '*boy*' and '*boys*' never occurred in the context of a noun phrase conjunction (e.g. both '*boys and men*' and '*John and boy*' were disallowed in the training corpus). Given these constraints we can test the net on

known words in novel syntactic positions as required by our definition of strong generalization and Hadley's notion of strong systematicity.[3]

The simple recurrent network employed in our simulations is a standard feedforward network equipped with an extra layer of so-called context units to which the activation of the hidden unit layer at time t is copied over and used as additional input at $t + 1$ (Elman, 1988, 1989, 1990, 1991a). We trained this net using incremental memory learning as proposed by Elman (1991b), providing the net with a memory window which 'grows' as training progresses (see Elman, 1993, for a discussion of the cognitive plausibility of this training regime). First, the net was trained for 12 epochs, resetting the context units randomly after every three or four words. The training set was then discarded, and the net trained for three consecutive periods of 5 epochs on separate training sets and with the memory window growing from 4–5 words to 6–7 words. Finally, the net was trained for 5 epochs on a fifth training set, this time without any memory limitations.[4]

In the remaining part of this section we will report the network's failure to exhibit strong generalization in genitive context and its success in obtaining strong generalization in the context of noun phrase conjunctions.

2.1 Limited Generalization in Genitive Context

Recall that neither 'girl' nor 'girls' has occurred in a genitive context in any of the training sets. Figure 5 illustrates the behavior of the net when processing the sentence 'Mary's girls run' in which the known word 'girls' occupies the novel syntactic position constituted by the genitive context (and the control sentence 'Mary's cats run').[5]

In (a) having received 'Mary . . .' as input, the net correctly predicts that the next word will either be a singular verb, a preposition, 'who', 'and', or a singular genitive marker. Next, the net expects a noun when given the singular genitive marker 'Mary's . . .' in (b). However, as can be seen from (e), which shows the activation of all the words in the noun category, the net neither predicts 'girl' nor 'girls' following a genitive marker. A similar pattern is found in (c) where the net expects a plural verb after 'Mary's girls . . .', but only provides the plural genitive marker with a small amount of activation (compared with the control sentence). The lack of generalization observed in both (c) and (e) indicates that the

---

[3] Hadley (personal communication) has acknowledged both test cases as possible *single* instances of strong sytematicity; though these instances may not be sufficient to warrant the *general* ascription of strong systematicity to the net as a whole.

[4] For more details about the grammar, the training regime, and additional simulation results, see Christiansen (in preparation).

[5] In Figure 5 (a)–(d) and 6 (a)–(j), **s-N** refers to proper/singular nouns, **p-N** to plural nouns, **s-V** to singular verbs, **p-V** to plural verbs, **prep** to prepositions, **wh** to *who*, **conj** to *and*, **s-g** to singular genitive marker, **p-g** to plural genitive marker, **eos** to end of sentence marker, and **misc** to *that* and the nouns used in the prepositions.

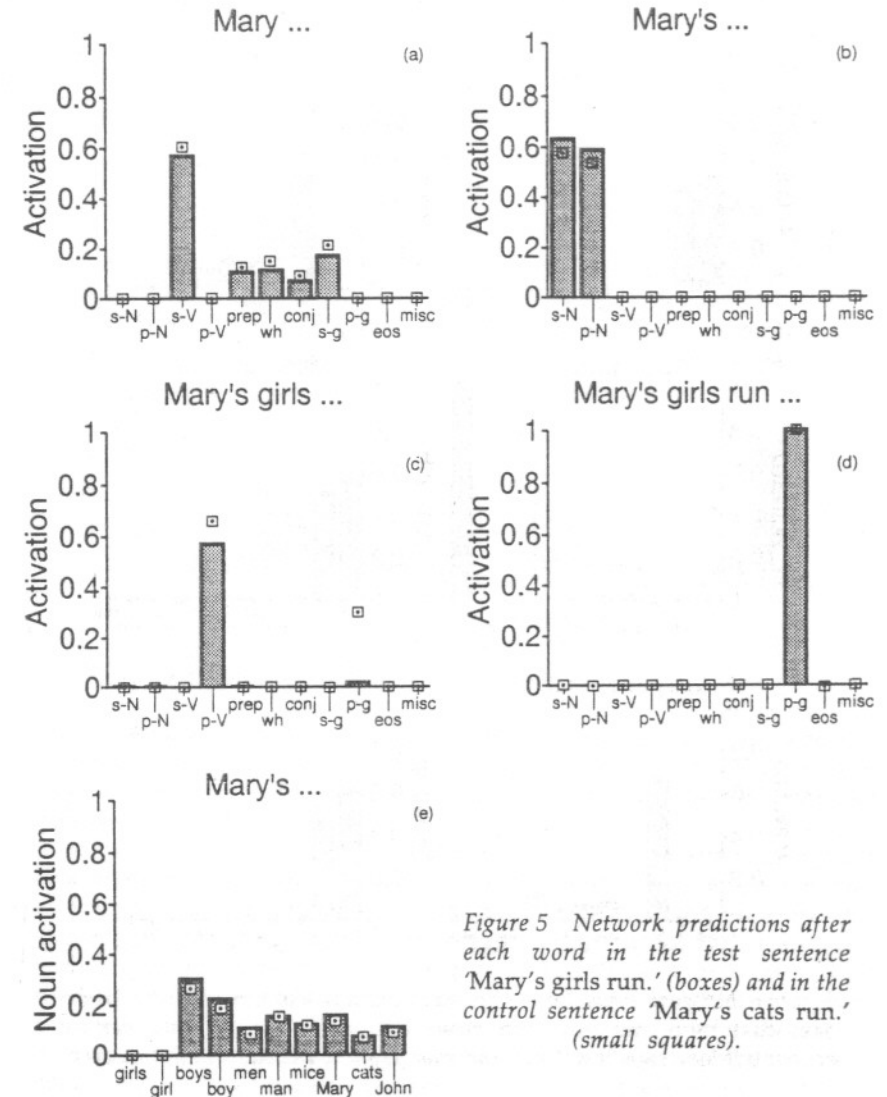*Figure 5   Network predictions after each word in the test sentence 'Mary's girls run.' (boxes) and in the control sentence 'Mary's cats run.' (small squares).*

net is not able to strongly generalize in genitive contexts. Notice that the net nonetheless is able to continue making correct predictions as shown by the high activation of the end of sentence marker after 'Mary's girls run . . .' in (d). Moreover, the fact that the net does activate the plural genitive marker in (c)—albeit by a very small amount—suggests that the net might be able to learn to strongly generalize in genitive contexts if a
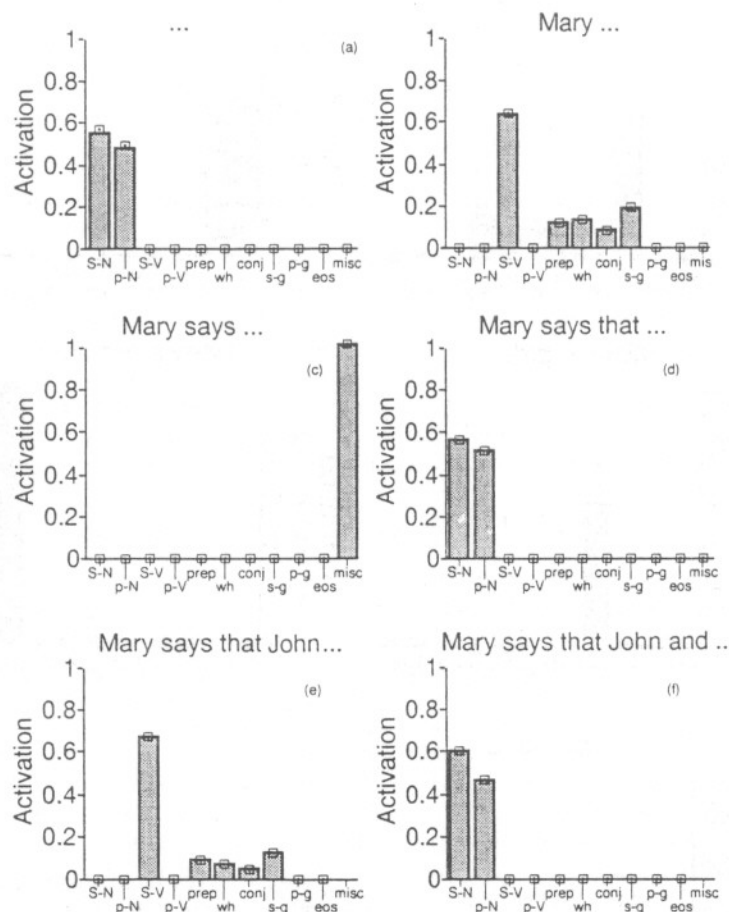
Figure 6 Network predictions after each word in the test sentence 'Mary says that John and boy from town eat.' *(boxes) and in the control sentence* 'Mary says that John and man from town eat.' *(small squares).*

different kind of representation is used or the details of the training are altered (we are currently pursuing both suggestions).

2.2 *Strong Generalization in Noun Phrase Conjunctions*

In contrast to the net's failure to strongly generalize in a genitive context, we shall now report the net's successful generalizations with respect to noun phrase conjunctions. Figure 6 illustrates network behavior during
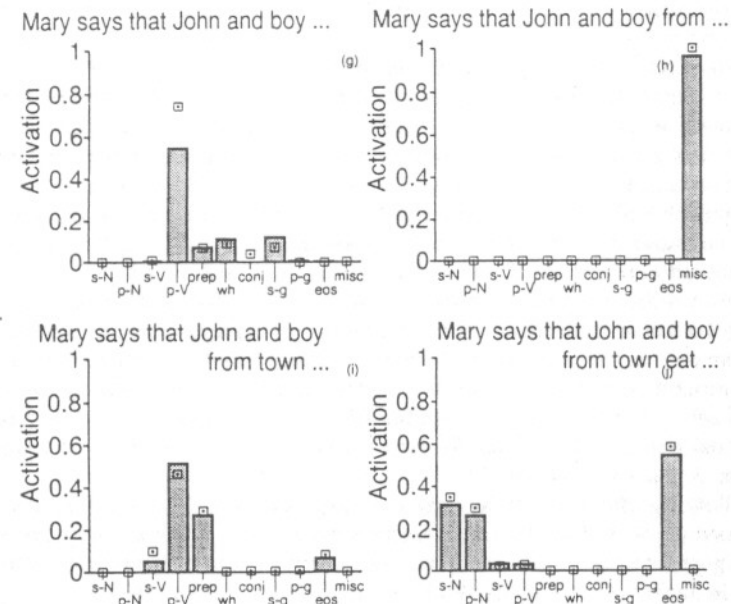
Figure 6 continued

the processing of the sentence '*Mary says that John and boy from town eat*' in which '*boy*' occurs in the novel syntactic context of a noun phrase conjunction (contrasted with the control sentence '*Mary says that John and man from town eat*').

At the beginning of a sentence the net expects to get either a singular or a plural noun (a). Given '*Mary . . .*' as in (b), the net predicts the next word to be either a singular verb, a preposition, '*who*', '*and*', or a singular genitive marker. In (c) the net has correctly categorized '*says*' as a clausal verb, expecting '*that*' to follow. A plural or a singular noun is rightly predicted in (d) given that the input so far is '*Mary says that . . .*'. '*John*' becomes a possible subject of the sentential clause, cueing the net to predict a singular verb, or a modification of '*John*' starting with a preposition, '*who*', '*and*', or a singular genitive marker as the next input. Having received '*and*' in (f), the net knows that a noun must follow (though the activations for both '*boy*' and '*boys*' are minimal compared with the other nouns). As we can see from (g), the net *is* able to correctly predict a plural verb following '*boy*' in '*Mary says that John and boy . . .*'. Despite only having seen a *singular* verb following '*boy*', the net is able to produce the strong generalization that a noun phrase conjunction always takes a *plural* verb. Admittedly, the activation of the plural verbs is not as high as in

the control sentence, but it is still significant and thus affords strong generalization (also in spite of the lack of activation of *'and'* compared with the control sentence). Notice also that the net predicts that a *singular* genitive marker might occur to modify *'boy'* (as well as a preposition or *'who'*). Since the next input is a preposition, the net predicts that the following word must be one of the nouns used with the prepositions (h). In (i) it is possible to detect two minor errors (singular verbs and end of sentence marker), but they are quite small (< 0.1). More importantly, the net gets the plural agreement right across the prepositional phrase. This is a considerable feat since it implies that the net is able to strongly generalize *across several words*. In particular, it shows that the net is not simply predicting plural verbs on the basis of having seen an *'and'* two items before, but has learned the grammatical regularities subserving noun phrase conjunctions.[6] Lastly, (j) demonstrates that not only is the net able to predict a correct end of sentence after *'Mary says that John and boy from town eat . . .'*, but it is also capable of predicting that *'eat'* might take an optional direct object (and the erroneous activations of both singular and plural verbs are very small).

Following the net's failure to strongly generalize in genitive contexts, we have shown that the net nevertheless is able to exhibit a quite robust strong generalization when faced with noun phrase conjunctions. Whether this instance of strong generalization is sufficient to endow the system with Hadley's notion of strong systematicity depends on whether two nouns out of a total number of ten nouns will count as a *'significant fraction of the vocabulary'* (Hadley, this issue, p. 251). Independent of the answer to this question, we agree with Hadley that human language learners presumably are able to strongly generalize in a number of different syntactic contexts. The genitive example illustrates that our net was not able to learn this more widespread (strong) generalization ability—yet such an ability might not be beyond simple recurrent networks as indicated by the net's successful (strong) generalization on noun phrase conjunctions.

3. Discussion

With our simulations we have taken a first step towards testing whether simple recurrent networks can meet our criterion of strong generalization. Our results show that this criterion is indeed difficult to meet, but suggest the possibility of future progress. We now turn to implications for the likely value of connectionist approaches to syntax acquisition.

A general objection to connectionist models of language learning is that

---

[6] Hadley (personal communication) has suggested that a context in which neither conjunct has previously occurred in a conjunction would constitute a better test of strong generalization. We followed this suggestion using the example of *'boy and boy'* and found the same pattern of results as mentioned above.

they use bottom-up statistical information about the training set, and that arguments from the poverty of the stimulus appear to rule out the feasibility of this approach. Whatever the merits of such arguments in general, in this context there is some evidence that simple statistical analysis of linguistic input may be sufficient to achieve strong generalization. Finch and Chater (1992, 1993; see also Maratsos, 1988) have shown that distributional statistics can be used to cluster lexical items into categories which have distinct primary syntactic classes, using a corpus consisting of, for example, forty million words from Internet newsgroups. In an attempt to study the learning task facing the child more closely, Redington, Chater and Finch (1993) applied this method to parent's child-directed speech from the CHILDES corpus (MacWhinney and Snow, 1985), again achieving good results. Redington et al. also studied how the syntactic category of a nonsense word could be derived from a single occurrence of that word in the training corpus. This corresponds to the set-up in the developmental studies that Hadley discusses (e.g. Gropen, Pinker, Hollander, Goldberg and Wilson, 1989; Pinker, Lebeaux and Frost, 1987). By comparing the single training context with the mean contexts associated with other items (which had been categorized by distributional analysis), Redington et al. found that classifications of syntactic categories could be made significantly above chance, at least for nouns and verbs. For example, given the sentence *'this is a wug'*, the novel word *'wug'* is successfully classified as a noun. These results imply that, for instance, a word which has only been encountered in object position can be recognized as a noun, and thus used in any legal noun position. This distributional analysis does not solve the whole problem that a connectionist system faces—learning not just lexical categories, but also the underlying regularities governing how those categories can be combined. Nonetheless, these results suggest that learning abstract syntactic categories from scratch, which is the key to strong generalization, may be possible using purely bottom-up information.

It should be noted that achieving strong generalization is not only a problem for connectionist approaches to syntax acquisition. For example, most symbolic approaches to language acquisition do not achieve strong generalization. For example, consider the model of language acquisition developed by Berwick and Weinberg (1984). In this model, the architecture of a Marcus-style parser (Marcus, 1980) is built in and provided with certain grammatical information, concerning, for example, X-bar theory (Jackendoff, 1977). However, the initial state contains no grammatical rules. Learning proceeds by attempting to parse input sentences; when parsing fails a new rule is added to the grammar so that parsing is made possible. Aside from the range of difficulties that this kind of model faces (for example, performance is catastrophically impaired if some input is ungrammatical), it does not begin to address strong generalization. This is because it does not *learn* the categories of new words from experience— instead, it is provided with information about the syntactic category of each word, including the argument structure of verbs, how these words

fit into larger syntactic structures, and so on. This means, for example, that when it encounters a new word, say, '*cat*', in a subject position, it is told that '*cat*' is a noun, making it trivial to process '*cat*' in an object position. Only when symbolic models are able to learn from unlabeled sequences of words will it be possible to assess the degree to which such models can exhibit strong generalization. The question of strong generalization is therefore just as pressing for symbolic approaches as for connectionist approaches to learning language.

In conclusion, Hadley has posed an important challenge to advocates of connectionist models of language. We have attempted to build on this work in order to set a clear objective for connectionist research on acquisition of linguistic structure. Our simulations indicate that this goal is not easily met, but nonetheless suggest that progress is possible. Only by properly addressing the issue of generalization can connectionist—and symbolic—models provide compelling accounts of human language acquisition.

Department of Philosophy
Washington University
St. Louis, MO 63130-4899
USA

Department of Experimental Psychology
University of Oxford
Oxford OX1 3UD
UK

References

Berwick, R.C. and Weinberg, A.S. 1984: *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition.* Cambridge, MA.: MIT Press.

Chater, N. 1989: Learning to Respond to Structures in Time. Technical Report RIPRREP/1000/62/89. Research Initiative in Pattern Recognition, St Andrews Road, Malvern, Worcs.

Chater, N. and Conkey, P. 1992: Finding Linguistic Structure with Recurrent Neural Networks. *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society.* Bloomington, IN.

Christiansen, M. 1992: The (Non)Necessity of Recursion in Natural Language Processing. *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society.* Bloomington, IN.

Christiansen, M. In preparation: Connectionism, Learning and Linguistic Structure.

Christiansen, M. and Chater, N. In preparation: Natural Language Recursion and Recurrent Neural Networks.

Elman, J.L. 1988: Finding Structure in Time. CRL Tech Report 8801, Center for Research in Language, University of California, San Diego.

Elman, J.L. 1989: Representation and Structure in Connectionist Models. CRL Tech Report 8903, Center for Research in Language, University of California, San Diego.

Elman, J.L. 1990: Finding Structure in Time. *Cognitive Science*, 14, 179–211.

Elman, J.L. 1991a: Distributed Representation, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7, 195–255.

Elman, J.L. 1991b: Incremental Learning, or The Importance of Starting Small. *Proceedings from the Thirteenth Annual Conference of the Cognitive Science Society.* Chicago, IL.

Elman, J.L. 1993: Learning and Development in Neural Networks: the Importance of Starting Small. *Cognition*, 48, 71–99.

Finch, S. and Chater, N. 1992: Bootstrapping Syntactic Categories by Unsupervised Learning. *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society.* Bloomington, IN.

Finch, S. and Chater, N. 1993: Learning Syntactic Categories: A Statistical Approach. In M. Oaksford and G.D.A. Brown (eds), *Neurodynamics and Psychology.* New York: Academic Press.

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

Jackendoff, R. 1977: $\bar{X}$ *Syntax: A Study of Phrase Structure.* Cambridge, MA.: MIT Press.

MacWhinney, B. and Snow, C. 1985: The Child Language Data Exchange System. *Journal of Child Language*, 12, 271–95.

Maratsos, M. 1988: The Acquisition of Formal Word Classes. In Y. Levy, I.M. Schlesinger and M.D.S. Braine (eds), *Categories and Processes in Language Acquisition.* Hillsdale, NJ.: Lawrence Erlbaum Associates.

Marcus, M. 1980: *A Theory of Syntactic Recognition for Natural Language.* Cambridge, MA.: MIT Press,

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

Redington, M., Chater, N. and M., Finch, S. 1993: Distributional Information and the Acquisition of Linguistic Categories: A Statistical Approach. *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society.* Boulder, CO.