C E N T E R   F O R   R E S E A R C H   I N   L A N G U A G E

Marrch 1991 **Vol. 5, No. 2**

The newsletter of the Center for Research in Language, University of California, San Diego, La Jolla CA 92093. *(619) 534-2536*; electronic mail: *crl @ crl.ucsd.edu*

CONTENTS

EDITOR'S NOTE

This newsletter is produced and distributed by the **CENTER FOR RESEARCH IN LANGUAGE**, a research center at the University of California, San Diego that unites the efforts of fields such as Cognitive Science, Linguistics, Psychology, Computer Science, Sociology, and Philosophy, all who share an interest in language. We feature papers related to language and cognition (1-10 pages, sent via email) and welcome response from friends and colleagues at UCSD as well as other institutions. Please forward correspondence to:

Jill Weckerly, Editor
Center for Research in Language, C-026
University of California, San Diego 92093
Telephone: (619) 534-2536
Email: *crl @ crl.ucsd.edu*

SUBSCRIPTION INFORMATION

If you are currently receiving a hardcopy of the newsletter only and have an email address to which the newsletter can be sent, please forward that information to CRL.

Can we send you the email version only to save printing and mailing costs? If you require a hardcopy in addition, please request it and we will be happy to send you one.

If you require the unformatted nroff version of the newsletter, please request it from CRL after you have received the current regular formatted version.

If you know of others who would be interested in receiving the newsletter, please forward the email or postal mailing address to CRL. Thank you.

BACK ISSUES

Back issues of this newsletter are available from **CRL** in hard copy as well as soft copy form. Papers featured in previous issues include the following:

The Cognitive Perspective
Ronald W. Langacker
Department of Linguistics, UCSD
vol. 1, no. 3, February 1987

Toward Connectionist Semantics
Garrison Cottrell
Institute for Cognitive Science, UCSD
vol. 1, no. 4, May 1987

Dimensions of Ambiguity
Peter Norvig
Computer Science, UC Berkeley
vol. 1, no. 6, July 1987

Where is Chomsky's Bottleneck?
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 1, no. 7, September 1987

Transitivity and the Lexicon
Sally Rice
Department of Linguistics, UCSD
vol. 2, no. 2, December 1987

Formal Semantics, Pragmatics,
and Situated Meaning
Aaron Cicourel
Department of Sociology, UCSD
vol. 2, no. 3, January 1988

Rules and Regularities in the
Acquisition of the English Past Tense
Virginia Marchman
Department of Psychology, UC Berkeley
vol. 2, no. 4, April 1988

A Geometric Conception of Grammar
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 2, no. 5, June 1988

Harris and the Reality of Language
S.-Y. Kuroda
Department of Linguistics, UCSD
vol. 3, no. 1, September 1988

A Connectionist Perspective on Prosodic Structure
Mary Hare, Dept. of Linguistics, UCSD
David Corina, Dept. of Psychology, UCSD
G.W. Cottrell, Dept. of Computer Science, UCSD
vol. 3, no. 2, November, 1988

Competent Scientist Meets the Empiricist Mind
Valerie Walker
Department of Philosophy, UCSD
vol. 3, no. 4, April, 1988

Foreign Policy By Metaphor
Paul Chilton, University of Warwick, U.K.
George Lakoff, UC Berkeley
vol. 3, no. 5, June, 1989

Guarani Possessive Constructions
Maura Velazquez
Department of Linguistics, UCSD
vol. 3, no. 6, July, 1989

Anticipatory Coarticulation and Aphasia: Implications
for Connectionist Models of Speech Production
William F. Katz
Department of Psychiatry, UCSD
vol. 4, no. 1, December, 1989

Toward a Connectionist Representation
of Grammatical Knowledge
Catherine L. Harris
Department of Cognitive Science, UCSD
vol. 4, no. 2, January, 1990

The Rule Controversy: A Cognitive Grammar Perspective
Ronald W. Langacker
Department of Linguistics, UCSD
vol. 4, no. 3, June 1990

Language and the Primate Brain
Martin I. Sereno
Department of Cognitive Science, UCSD
vol.4, no. 4, August, 1990

The Segmentation Problem in Early Language Acquisition
Kim Plunkett
University of Aarhus, Denmark
vol.5, no. 1, November, 1990

NEO-STRUCTURALISM : a commentary on the correlations between the work of Zelig Harris and Jeffrey Elman

Peter A. Bensch
Department of Computer Science & Engineering
University Of California, San Diego

ABSTRACT

Elman's current connectionist models are designed to extract significant regularities from "streams" of input data. A key component of this work is the inclusion of time. That is, these networks use both the current input and the previous network state to determine their output.

This inclusion of time has allowed his networks to recognize regularities that were beyond the reach of earlier network designs. Significantly, his networks' outputs very closely follow the predictions of Harris [HARRIS82]. (Harris is one of the last remaining practitioners of pre-Chomskyan structuralism.) The Chomskyan revolution was to some extent precipitated by the lack of sufficient computational tools to meet the goals of linguistic structuralism. Chomsky proposed that the structuralist program of inducing general principles from empirical data would never succeed. As part of his revolution, he advocated a research program based on deduction from general principles to empirical data.

With the emergence of the computational tools being developed by Elman, structuralism may again become a viable research program. Further support for this conjecture is provided by the continuing problems encountered by linguists attempting to deduce empirical data from base principles. Thus, a connectionist revolution seems to be emerging. And, this revolution may be fittingly called "neo-structuralism."

1. INTRODUCTION

In the past, connectionist models have concentrated on developing a cause and effect relationship between the model's inputs and its outputs. When a model was successful, analysis would reveal that it had extracted key components (features) from the inputs that were good predictors of the output. This program was successful as long as the inputs were causal determiners of the outputs.

The program was successfully challenged by Pinker and Prince [PINKER88]. In that paper, they identified many problems in the Rumelhart and McClelland [RUMEL86a]model of the acquisition of verbal past-tense forms. Among those problems was the fact that the root form of the verb was not a sufficient predictor of its past-tense form. Much recent connectionist work has been directed toward overcoming the problems identified by Pinker and Prince.

As part of this recent work, Elman has been extending the work of Jordan [JORDAN86]. This work includes time as an element in the causal relationship between inputs and outputs (that is, the connectionist network uses both the input and its own previous state to determine its output). In particular, Elman has found that having the network predict the next input has been a successful method for identifying significant regularities in the input data. In [ELMAN90], he reports on a model that can identify word boundaries when the input is a stream of letters that spell out the words (with no indication of where one word ends and the next begins). Another model can identify word categories when given a stream of input consisting of words that form sentences (with no indication of where one sentence ends and the next begins). In both of these models, the input is not a sufficient predictor of the output. However, the inclusion of the network's state as part of the input allows the network to successfully extract relevant statistical properties from the input. Although this does not directly address the prediction of the verbal past-tense form, it does open a new avenue of research that might aid in solving the problem.

Elman's work is reminiscent of the work performed by the pre-Chomskyan linguists. These "structuralists" believed that the correct way to study language was to look for regularities in actual usage. That is, they believed that word categories, grammar, etc. could be identified by direct observation of the "word streams" produced by native speakers of the language. Unfortunately, the computational tools available to the linguists of the early-1950's were not sufficient to make this program successful.

In the mid-1950's, Chomsky broke from the "structuralist" tradition, starting his revolution in linguistics. His break included a key change in methodology. He believed that actual language was an aberration. It contained far too much "noise" to ever allow the induction of general principles. Instead, he postulated two levels of language: performance and competence. Whereas the "structuralists" were committed to the study of performance, he believed that there was a Platonic, internal language competence. He felt that this internal language must be very orderly, free from the noise present in language user's "word streams." Thus, Chomsky's research program involved: (1) postulating the principles underlying the internal language, (2) deducing grammar, etc. of the internal language from these basic language principles, and (3) describing how the "noisy" language performance arises from the internal language competence. Thus, Chomsky's revolution involved a very significant shift in methodology. Post-Chomsky linguists perform deduction from basic principles to the empirical language data. Pre-Chomsky linguists performed induction from the empirical language data to basic principles.

I believe that Elman's recent connectionist models are built in the pre-Chomskyan tradition. His goal is to have his networks discover linguistic properties from empirical data. In a recent critique of the connectionist research program, Fodor and Pylyshyn [FODOR88] argued forcefully that connectionism was only valid as a computational tool being used to implement post-Chomsky linguistic theories [see NOTE 1]. This conviction is firmly grounded in their commitment to a deductive methodology. They can see no "revolution" in the connectionist program. However, when viewed the way I am proposing, there is a very definite connectionist revolution -- or, better, connectionist counter-revolution. If Elman is successful in his current research program, he will indeed have created a powerful computational tool. It may be just the tool that the "structuralists" of the early-1950's lacked. If so, we may be witnessing the origin of a neo-structuralism.

In the following, I will discuss the implications of a neo-structuralist revolution. Then I will discuss in detail two of Elman's recent models, and contrast them with the current work of Zelig Harris -- a linguist that has continued to follow the structuralist methodology.

2. THE CONNECTIONIST REVOLUTION

Among the problems that will face the modellers following Elman's research program will be the selection of the appropriate kind of data to use in training their models. There are two possible choices for training data: (1) actual examples from a human language, or (2) an artificial sample generated from a standard grammar. The problem here is that using the latter presupposes the deductive base principles underlying the artificial grammar.

Assume a connectionist modeller begins an experiment by generating training sentences from a grammar. Then these sentences are used to train the model. Finally, the model is tested by using actual sentences from the language. This seems like a perfectly valid research program, but it is fatally flawed. The problem is that the model will learn the formal system, not the language. If the model performs perfectly well in the test sentences given it, this says nothing about the validity of the model as a whole. At any time, we might encounter a series of sentences that the model could not process correctly. The only value in this model would be in refuting the formal system under study. Since this kind of negative evidence could easily be generated without the construction of such a model, this seems to be a fruitless enterprise.

To be fair, I should note possible alternative motives for creating such a model. For classical (i.e., Chomskyan) linguists, such a model would show that the formal system under study could in fact be created using "neuron-like" machinery. This could aid in establishing the psychological reality of the model. Further, such a simulation could be used to show that the behavioral consequences of the formal system parallel human behavior with the target language. But, again, the object of study is the formal system, not the language itself. On the other hand, this kind of model could be used by a connectionist to show that his/her model can do at least as much as a classical model can.

The other alternative is to use actual language sentences to train the connectionist model. The problem here is that the classical Chomskyan linguists have nothing to offer our modeller in this case. One of their key tenets is that language principles are not induced from the empirical language data. Thus, our modeller is left without any theorectical underpinnings to aid his/her research. He/she will be like Thomas Kuhn's pre-paradigm scientists:

> In the absence of a paradigm or some candidate for paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. As a result, early fact-gathering is usual-

> ly restricted to the wealth of data that lie ready to hand. The resulting pool of facts contains those accessible to casual observation and experiment together with some of the more esoteric data retrievable from established crafts ...
>
> [KUHN, pp.15]

In this case the craft will be connectionist modelling, and the data that has emerged thus far from sentence processing models has been diverse and difficult to analyze.

But, we must recall, that linguistics is not a pre-paradigm science. It is, in fact, a science with its paradigm -- generative grammar -- in crisis. But paradigms under attack can prove to be quite resilient. Kuhn says the following about scientists struggling through such a crisis:

> Though they may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis. They do not, that is, treat anomalies as counterinstances, though in the vocabulary of philosophy of science that is what they are. ... [O]nce it has achieved the status of paradigm, a scientific theory is declared invalid only if an alternative candidate is available to take its place. ... The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgement leading to that decision involves the comparison of both paradigms with nature and with each other.
>
> [KUHN, pg.77]

If the generative grammar paradigm has no assistance for our connectionist modellers, is there any other source that might help? The answer is yes. As Kuhn has noted, when paradigms begin to lose their dominance, the research they guide increasingly resembles "that conducted under the competing schools of the pre-paradigm period" [KUHN, pg.72]. The school of linguistics that preceded the Chomskyan paradigm was post-Bloomfieldian structuralism, and one of its foremost practitioners was Zelig Harris. In fact, Harris has continued to practice and refine structural linguistics throughout the Chomskyan revolution. S.-Y. Kuroda notes:

> the difference between Harris and Chomsky turns on the notion of grammar. Harris was one of the foremost methodologists in post-Bloomfieldian taxonomic structuralism; he brought it to a completion by his work Methods in Structural Linguistics in 1947. Harris attempted to extend the taxonomic methodology of descriptive linguistics to discourse analysis around 1950, but by 1960 he had virtually returned to the study of grammar by developing [his] transformational theory, without explicitly dissociating himself from his past methodological stance. Chomsky, in the meantime, abandoned taxonomic methods of structural linguistics in the early 50's and launched into the construction of the theory of transformational generative grammar under a "realist" and pyschological interpretation of linguistic theory.
>
> [KURODA, pg.45]

Expounding on the differences between Harris and Chomsky, Kuroda says

> Harris's [transformational] theory is directed to the structure of correspondence that underlies the syntactic design of language. ... Correspondence and derivation are two dynamic forces that shape the formal design of human language, and it is a major task imposed on linguistic theory how to determine the sphere of influence of these contending forces. Harris' transformational theory took the form it did to respond primarily to the former, and Chomsky's initial formulation of transformational generative grammar, to the latter. The later development of transformational generative grammar may to a large measure be looked upon as testimony to the linguist's response to a tension produced by two contending forces.
>
> [KURODA, pg.6]

In further examining the history of generative grammar, Kuroda notes:

> Chomsky is reported to have ... expressed the opinion that "the history of transformational grammar would have been more 'rational' if generative semantics had been the original position ..." ... [A] development from generative semantics through the Standard Theory and then to the Government and Binding Theory is easy to imagine as a rational history of transformational grammar ... If what interests us is a conceivable ideal history, ... one might be able to imagine a path from Harris' ... conception of transformational theory to the present [i.e., Government and Binding Theory] and to the future, without going through the idea of transformational generative grammar ...
>
> [KURODA, pg.47]

Thus, it appears that Chomsky's theory,

emphasizing derivation, and Harris' theory, emphasizing correspondence, are two possible trails leading to the same end. What is important here for connectionist modellers is that Harris' theory gets to the common goal via the study of actual language performance. Thus, Harris' theory may provide connectionist modellers the appropriate guidance to be successful in developing their grammar models. Further, Harris' theory provides specific guidance as to what types of internal representation might be expected to emerge from these models. This should guide the modellers as they attempt to analyze their models' performance.

Below, I will examine two specific Elman models. Both of these models take sentences, represented by a stream of words, as their input. The "simulations address problems in the distinction between type and token, the representation of lexical categories, and the representation of grammatical structure." [ELMAN89, pg.1]

At the core of both simulations is the way words interact in the sentences of a language. Harris' theory is also built on a foundation of word interactions. Harris' theory postulates the emergence of "grammar-like" behavior from these low-level interactions. This too, appears to be happening in the models under review. I will have more to say about each model in turn.

3. LEXICAL CATEGORY STRUCTURE

In his first model, Elman sought to demonstrate that "a network could learn the lexical category structure which was implicit in a language corpus." [ELMAN89, pg.3] A key assumption behind this model was:

> One of the consequences of lexical category structure is word order. Not all classes of words may appear in any position. Furthermore, certain classes of words ... tend to cooccur with other words.
>
> [ELMAN89, pg.3]

The network was trained "to take successive words from the input stream and to predict the subsequent word" [ELMAN89, pg.4]. After being trained on 6 cycles through 10,000 two- and three-word sentences, the network's internal representations were examined. The hidden unit activations were averaged over all occurrences for each word in the lexicon. Then, these "mean vectors" were analyzed using "hierarchical clustering analysis." The resulting similarity structure shows a grouping of the words by the traditional lexical categories of verb and noun. The verbs are further divided by their argument requirements. The nouns are divided into animates and inanimates. And each of these categories is further divided into groups based on the set of verb argument roles they can fill.

Elman summarizes the model's performance as follows:

> The network is not able to predict the precise order of specific words, but it recognizes that (in this corpus) there is a class of inputs (viz., verbs) which typically follow other inputs (viz., nouns). This knowledge of class behavior is quite detailed; from the fact that there is a class of items which always precedes chase, break, and smash, it infers a category of large animals (or possibly, aggressors).
>
> [ELMAN89, pg.7]

In Harris' theory, the ability of words to enter together into a sentence is based on their likelihood to co-occur. This likelihood is a statistical relationship between words that can be observed over time. In Harris' theory all words are operators. Operators can take zero or more arguments, with the operator's first argument always preceding it. Words that can start sentences form a priviledged class as operators with no argument predecessors. These "null" operators are labelled N. All other operators are labelled O(x,y,...), where x, y, ... identify the class of arguments that will co-occur with the operator. X and y are always chosen from n (for "null" operators) or o (for all other kinds of operators). Thus, as Harris is gathering co-occurrence statistics, he is also using the sequential ordering of words to determine operator classes.

When a corpus of language material has been analyzed in this way, Harris predicts that all words will fall into large operator classes. In this particular example, the nouns will be classified as type N operators, and the verbs will be classified as either a type O(n) operator -- intransitive -- or as a type O(n,n) operator -- transitive. Note that Elman identified a third class of verb with an optional direct object. Harris would eliminate this class by making the verb transitive and claiming that its object was "reduced" to zero. In fact, he would not have included any of the sentences with the missing direct objects in his "base" corpus on strictly theoretical grounds. Thus, this would be a case where having an operational paradigm to follow would have influenced the data selected to train the model.

(Note that Harris would not object to having a missing direct object in the test set for the model. He would predict that the model would partially activate all of the possible objects when the verb is presented. Then, when no object occurred, the model should have a "very likely" object highly activated. This object could be at the word level, or it could be at a higher "word group" level. In both cases, the object would be providing little or no information to the sentence and would be a candidate for reduction -- see TRANSFORMATIONS below.)

The main goal of Harris' analysis is to determine co-occurence "likelihoods" between words in the target lexicon. A further subdivision of the operator classes will occur based on the similarity between the co-occurence sets associated with individual words. From the fact that there is a class of items within the N operator class which always acts as the first argument for the O(n,n) operators chase, break, and smash, it follows that a subdivision of "large animals" will occur in the N operator class.

Note that in Harris' theory these co-occurence sets are "fuzzy". They are dynamic -- subject to change as the language users vary the meaning of their words. Thus, at any point in time, a word's co-occurence set would reflect all previous experience with that word over time. In other words, the co-occurence set is a direct analog of Elman's "mean vector" of hidden unit activations. However, the "lexical" operator classes that words belong to will remain constant over time. This again is based on theoretical considerations. Harris attempts to restrict all words to one operator class, leaving any appearance of membership in multiple classes to be explained by grammatical reductions.

4. TYPE-TOKEN DISTINCTIONS

Elman also clustered the hidden unit activation patterns for each word in the training data set. This "context-sensitive" clustering of hidden unit patterns created groupings similar to those obtained for the "mean vector" analysis.

> In this simulation, the context makes up an important part of the internal representation of a word. ... [I]t is literally the case that every occurrence of a lexical item has a separate internal representation.
> ... The fact that these are all tokens of the same type is not lost ... These tokens have representations which are extremely close in space -- closer to each other by far than to any other entity. Even more interesting is that the spatial organization within the token space is not random but reflects differences in context which are also found among tokens of other items. The tokens of boy which occur in subject position tend to cluster together, as distinct from the tokens of boy which occcur in object position. This distinction is marked in the same way for tokens of other nouns. Thus, the network has learned not only about types and tokens, and categories and category members; it also has learned a grammatical-role distinction which cuts across lexical items.
>
> [ELMAN89, pp.7-8]

Although Harris does not directly address this type-token distinction, he does address the emergence of grammatical-role from co-occuring words. The "fuzzy" sets of next words tend to establish grammatical-roles. In essence, the likelihood relationship between a word and its possible successors partitions the appropriate operator space in a very specific manner. In the context of a PDP schema model [RUMEL86b], each word will adjust the "goodness-of-fit" landscape for the next possible word. This distortion will place more likely words at very high points, and less likely words at lower points.

I believe that Elman's type-token distinction may well correspond to a word's adjusting positions in "likelihood" space based on the word(s) that preceded it. Note that subjects, which precede their verbs, would have a distinctly different position in "likelihood" space from objects, which follow the verb. Thus, it appears that Elman's type-token distinction is also consistent with Harris' language theory.

5. TRANSFORMATIONS

> Although lexical information plays an important role in language, it actually accounts for only a small range of facts. Words are processed in the contexts of other words; they inherit properties from the specific grammatical structure in which they occur.
>
> [ELMAN89, pg.8]

Up to now we have been looking at low-level relationships between Harris' word categories. But, at the next higher level, we can examine words which yield the same "fuzzy" sets for next words. These words can be considered "equivalent" to the extent that their word groups establish the same "context" for the next word. Thus, we can identify groups of subjects that are associated with the same "likelihood" space of verbs. Or, groups of verbs that are associated with the same "likelihood" space of objects. This equivalence relation between words, when correlated with the lexical definition of the words, can be used to identify word sequences that are paraphrases of each other. In fact, a necessary condition for Harris to consider two word sequences "paraphrastic" to each other is that they have the same next word "likelihood" space.

Harris considers word groups belonging to the same paraphrastic equivalence class to be related by linguistic transformations. He attempts to locate the core of a language by finding the one "kernel" word group for each class. This "kernel" word group must generate the whole class using as few transformations as possible. The analysis at this level will yield a set of transformation domains. Such a domain includes the words that terminate each word group on which the transformation can act. It should be noted that most of the transformations will be "reductions" -- that is, the elimination of redundant or low-information words. Further, these reductions are based on the "likelihood properties" of the component words of each word group.

A key point here is the need to include a "semantic" component to guide the network's search for transformations. Elman voices a similar sentiment:

> The network has no information available which would "ground" the structural information in the real world. In this respect, the network has much less information to work with than is available to real language learners. In a more realistic model of acquisition, one might imagine that the utterance provides one source of information about the nature of lexical categories; the world itself provides another source. One might model this by embedding the "linguistic" task in an environment; the network would have the dual task of extracting structural information contained in the utterance, and structural information about the environment. Lexical meaning would grow out of the associations of these two types of input.
>
> [ELMAN90, pg.201]

The appeal of a Chomskyan-style formal system is the ability to isolate syntax from semantics. However, as Chomsky has said, such a formal system cannot emerge by induction from the actual sentences of the language. Harris offers a theory that will allow grammar to arise from the actual linguistic data, but it requires the mixing of semantics with syntax.

6. GRAMMATICAL STRUCTURE

In the Elman model mentioned above, the corpus of sentences is generated by a very simple sentence generator. It had a set of simple two- and three-word sentence "templates" that it randomly filled with words from the lexicon. This corpus was so constrained that it would easily satisfy Harris' criterion for forming a sublanguage [HARRIS89]. A sublanguage is a very restricted subset of the language as a whole. The key restriction is that the words assigned to the sublanguage only have a "standard" usage. That is, the co-occurrence patterns are sufficiently regular that sentence "formulas" can be identified for those words. These sentence formulas perform the function of a grammar in Harris' theory of language.

Thus, Harris would predict that Elman should be able to identify specific sentence formulas in his model. Elman does not address this point in either of his two papers [ELMAN89,90] covering the first model. However, analysis using principle components (see below) identifies patterns in the hidden units that might qualify as sentence formulas.

Elman's second sentence processing model was a giant step beyond the one mentioned above. It's primary goal was to investigate a connectionist model's representation of grammatical structures. To pursue that goal, he set up the following training data set:

> The stimuli in this simulation were based on a lexicon of 23 items. These included 8 nouns, 12 verbs, the relative pronoun who, and an end of sentence indicator, ".". Each item was represented by a randomly assigned 26-bit vector in which a single bit was set to 1 (3 bits were reserved for another purpose). A phrase structure grammar ... was used to generate sentences.
>
> [ELMAN89, pg.9]

As mentioned above, when a modeller uses grammar generated sentences to train his/her model, the subject of study becomes the grammar, not the language. In this case, the goal was to show that a connection model could implement the rather complex system represented by the grammar. In particular, the grammar allowed the nesting of relative clauses. This made the tasks of subject/verb agreement and verb argument determination far more complex. Any of the words filling these roles might be separated from their companion word by one or more relative clauses. Elman found:

> the network was unable to learn the task when given the full range of complex data from the beginning of training. However, when the network was permitted to focus on the simpler data first, it was able to learn the task quickly and then move on successfully to more complex patterns. The important aspect to this was that the earlier training constrained later learning in a useful way; the early training forced the network to focus on canonical versions of the problems which apparently created a good basis for then solving the more difficult forms of the same problem.
> [ELMAN89, pp.11-12]

Since we are not talking about a corpus from actual language here, Harris is not really applicable. However, the idea that the model would learn the simpler patterns first is compatible with Harris. He would hold that the complex sentences would be "paraphrastically" equivalent to simpler sentences in the "kernel" language. Since the simpler sentences are all "kernel" sentences themselves, it would be easier to learn them. Learning a complex sentence would require the language learner to: (1) first acquire the "kernel" sentences that would be considered equivalent to the complex sentence, and (2) then learn the transformation(s) that relate the "kernel" and complex sentences [see NOTE 2]. If the corpus was from actual language, the frequency of occurrence of complex sentences would probably be diminished enough so that the task could be accomplished without resorting to the "staged learning" strategy used by Elman.

The corpus used for training this model was sufficiently simple so that the network could, in fact, learn its regularities without resorting to transformations. Thus, Harris would anticipate that sentence formulas should be stored within the statistical information coded by the hidden units. Elman also aniticipated that the grammatical structure must be coded in the hidden units. Since the cluster analysis only yielded categorical information, it was necessary to devise a different analysis technique to look for the grammatical relations. The technique that located this information was principle component analysis.

> This involved passing the training set through the trained network (with weights frozen) and saving the hidden unit pattern produced in response to each new input. The covariance matrix of the set of hidden unit vectors is calculated, and then the eigenvectors for the covariance matrix are found. The eigenvectors are ordered by the magnitude of their eigenvalues, and are used as the new basis for describing the original hidden unit vectors. This new set of dimensions has the effect of giving a somewhat more localized description to the hidden unit patterns, because the new dimensions now correspond to the location of meaningful activity (defined in terms of variance) in the hyperspace. Furthermore, since the dimensions are ordered in terms of variance accounted for, we can now look at phase state portraits of selected dimensions, starting with those with the largest eigenvalues.
> [ELMAN89, pg.15]

In particular, Elman found that principal components 1 and 11 appear to identify the sentence formulas for the following test sentences:

(10a) boy chases boy .
(10b) boy chases boy who chases boy .
(10c) boy who chases boy chases boy .
(10d) boy chases boy who chases boy who chases boy .

[ELMAN89, pg.17]

> The trajectories through state space for these four sentences ... are shown in Figure 10 [pg.18]. Panel (10a) shows the basic pattern associated with what is in fact ows the basic pattern associated with what is [T]he matrix subject noun is in the lower left region of state space, the matrix verb appears above it and to of state space, the matrix verb appears above it and to the left, and the matrix object noun is near the upper middle region. ... The relative clause appears to involve a replication of this basic pattern, but displaced toward the left and moved slightly downward, relative to the matrix constituents. Moreover, the exact position of the relative clause elements indicates which of the matrix nouns are modified. ... This trajectory pattern was found for all sentences with the same grammatical form; the pattern is thus
> [ELMAN89,pp.17-18]

Thus, it appears that another of Harris' predictions is being fulfilled. It is possible to identify the underlying grammatical structure for a simple corpus by induction from the empirical data.

7. CONCLUSION

The correlations between Elman's and Harris' work seem to be quite strong. This implies that the models and analysis techniques that Elman has been developing might prove very useful for linguistic structuralists. Elman, himself, seems to have become a methodological structuralist. Given that Harris' work implies a realist position [KURODA], Elman may not be too upset being characterized as a neo-structuralist.

ACKNOWLEDGEMENT

I would like to thank Walt Savitch for many stimulating discussions concerning the subject matter of this paper. I would like to further thank Savitch for helpful critical comments on earlier drafts of this paper.

NOTES

1. The following comments by Fodor and Pylyshyn indicate their strong feeling that connectionism should be considered an implementation theory for classical cognitive theories.

many of the arguments for Connectionism are best construed as claiming that cognitive architecture is implemented in a certain kind of network (of abstract "units"). Understood this way, these arguments are neutral on the question of what the cognitive architecture is. ...

[FODOR88, pg.64]

the implementation, and all properties associated with the particular realization of the algorithm that the theorist happens to use in a particular case, is irrelevant to the psychological theory; only the algorithm and the representations on which it operates are intended as a psychological hypothesis. ...

Given this principled distinction between a model and its implementation, a theorist who is impressed by the virtues of Connectionism has the option of proposing PDP's as theories of implementation. But then

... these models are in principle neutral about the nature of cognitive processes. In fact, they might be viewed as advancing the goals of Classical information processing psychology by attempting to explain how the brain (or perhaps some idealized brain-like network) might realize the types of processes that conventional cognitive science has hypothesized.

[FODOR88, pg.65]

2. Harris says the following about relative clauses:

English has a set of pronounings from which are derived all the modifiers in the language -- attributive adjectives, relative clauses, adverbs and PN phrases, and also subordinate clauses. All of these originate in relative clauses. The relative clause is a "secondary" sentence S2 connected by [a] semicolon to a "primary" sentence S1, where a word in S2 is reduced to a wh-pronoun on the grounds that it is the same as a word (the 'host') in S1. The wh-pronouning is carried out primarily on a word that is first in S2 (in many cases because of its front positioning as in section 3.11 [pp.109-115: Bill spoke to John; John knew Bill well --> Bill spoke to John; Bill John knew well]), when S2 has interrupted S1 (3.13 [pp.116-117: Bill spoke to John; Bill John knew well --> Bill -- Bill John knew well -- spoke to John]) immediately after the host. In this situation the two words that are the same are most often next to each other as in Bill -- Bill John knew well -- spoke to John --> Bill, whom John knew well, spoke to John. Although some of the sentences with front positioning may seem uncomfortable when standing alone, they are natural when interrupting a primary sentence after the same word as they placed in front position.

[HARRIS82, pp.120-121]

REFERENCES

[ELMAN89] Elman,J.L. (1989). Representation and structure in connectionist models.*CRL Technical Report 8903*, Center for Research in Language, University of California, San Diego.

[ELMAN90] Elman,J.L. (1990) Finding structure in time. *Cognitive Science*, 14, 179-211.

[FODOR88] Fodor,J.A., & Pylyshyn,Z. (1988). Connectionism and cognitive architecture: A critical analysis. In S.Pinker & J.Mehler (Eds.). *Connections and Symbols*. Cambridge, MA: MIT Press.

[HARRIS82] Harris,Z. (1982).*A Grammar of English on Mathematical Principles*. New York,NY: John Wiley & Sons.

[HARRIS89] Harris,Z. (1989). *The Form of Information in Science: Analysis of an Immunology Sub- language*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

[JORDAN86] Jordan,M.I. (1986). Serial order: A parallel distributed processing approach. *ICS Report 8604*, Institute for Cognitive Science, University of California, San Diego.

[KURODA] Kuroda,S.-Y. (1989). Derivational and geometric conceptions of grammar: Reflections on Harris and Chomsky. (Unpublished manuscript).

[KUHN] Kuhn,T.S. (1970).*The Structure of Scientific Revolutions (2nd Ed.)*. Chicago,IL: The University of Chicago Press.

[PINKER88] Pinker,S., & Prince,A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S.Pinker & J.Mehler, op. cit.

[RUMEL86a] Rumelhart,D.E., & McClelland,J.L. (1986). On learning the past tenses of English verbs. Chapter 18, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. II)*. Cambridge,MA: MIT/Bradford.

[RUMEL86b] Rumelhart,D.E., Smolensky,P., McClelland,J.L., & Hinton,G.E. (1986). Schemata and sequential thought processes in PDP models. Chapter 14, *PDP (Vol. II).*