

## The (Non)Necessity of Recursion in Natural Language Processing

Morten H. Christiansen\*

Centre for Cognitive Science  
University of Edinburgh  
2 Buccleuch Place  
Edinburgh EH8 9LW  
Scotland, UK  
*morten@uk.ac.ed.cogsci*

### Abstract

The *prima facie* unbounded nature of natural language, contrasted with the finite character of our memory and computational resources, is often taken to warrant a *recursive* language processing mechanism. The widely held distinction between an idealized infinite grammatical competence and the actual finite natural language performance provides further support for a recursive processor. In this paper, I argue that it is only necessary to postulate a recursive language mechanism insofar as the competence/performance distinction is upheld. However, I provide reasons for eschewing the latter and suggest that only data regarding observable linguistic behaviour ought to be used when modelling the human language mechanism. A connectionist model of language processing—the simple recurrent network proposed by Elman—is discussed as an example of a non-recursive alternative and I conclude that the computational power of such models promises to be sufficient to account for natural language behaviour.

### Introduction

Is it necessary to postulate a recursive language mechanism in order to account for the apparently unbounded complexity and diversity of natural language (NL) behaviour, given the finite nature of the memory and computational resources that underly the human production of this behaviour? What seems to be needed in the first place is a mechanism which is able to generate, as well as parse, an infinite number of NL expressions using only finite means. Obviously, such a mechanism has to be of considerable computational power and, indeed, recursion provides a very elegant way of achieving this property. Consequently, recursion has been an intrinsic part of most accounts of NL behaviour—perhaps due to the essentially recursive character

of most linguistic theories of grammar.<sup>1</sup>

It is often noted that the *prima facie* existence of recursion in NL behaviour poses serious problems for connectionist approaches to NL processing (e.g., Fodor & Pylyshyn, 1988) since recursion—*qua* computational mechanism—is defined as being essentially symbolic. However, the existence of recursion in NL presupposes that the grammars of linguistic theory correspond to *real* mental structures, rather than mere structural *descriptions* of NL *per se*. Yet, there are no *a priori* reasons for assuming that the structure of the observable public language necessarily must dictate the form of our internal representations (van Gelder, 1990b). Still, many linguists and psychologists (e.g., Chomsky, 1986; Frazier & Fodor, 1978; Kimball, 1973; Pickering & Chater, 1992; Pulman, 1986) take grammars as corresponding to in-the-head representations that are manipulated by computational processes. But, since human NL behaviour is limited under normal circumstances, a distinction is typically made between the bounded observable performance and an infinite competence inherent in the internal grammar.

In what follows, I start off by arguing from a methodological perspective that the alleged distinction between linguistic competence and actual NL performance must be rejected if linguistic theories are to encompass representational claims regarding the human NL mechanism. Then, I drive a wedge between the (quasi-) recursive nature of NL, as described in most current linguistic theories, and the actual NL processing mechanism. In particular, I suggest that recursion is a *conceptual artifact* of the competence/performance distinction (C/PD), instead of a necessary characteristic of the underlying computational mechanism.<sup>2</sup> In this light, the prob-

<sup>1</sup>For example, in GB (e.g., Chomsky, 1981) the underlying principles of  $\bar{X}$ -theory are recursive, as are the ID-rules of GPSG (Gazdar *et al.*, 1985).

<sup>2</sup>I will therefore not discuss connectionist models of NL processing that merely simulate—or mirror—symbolic recursion. An example of such models is provided by McClelland & Kawamoto (1986) who apply

This research was made possible through award No. V910048 from the Danish Research Academy.

lem facing connectionist models of NL processing is *not* whether they can implement some kind of recursive mechanism, but whether they will be able to account for the (limited) *recursive structure* found in NL behaviour purely in terms of non-symbolic computation. I therefore consider a connectionist model—designed by Elman (1990, 1991)—which exhibits recursive behaviour without implementing a symbolic recursion mechanism, and conclude that such connectionist models provide a psychologically appealing way of modelling NL processing.

### The Competence/Performance Distinction

In most—if not all—linguistic theories of NL, recursion is unbounded. However, since the main source of data of modern linguistics implies intuitive grammaticality judgements (e.g., Horrocks, 1987), the fact has to be explained that the greater the length and complexity of utterances, the less sure people are of their respective judgements. To explain this phenomena, a distinction between an idealized infinite linguistic *competence* and a limited NL *performance* is made. The performance of a particular individual is limited by memory limitations, attention span, lack of concentration, etc. (e.g., Fodor & Pylyshyn, 1988; Horrocks, 1987).

This methodological separation of the infinite linguistic competence of a recursive grammar from the limited performance of observable NL behaviour has been strongly advocated by Chomsky:

One common fallacy is to assume that if some experimental result provides counterevidence to a theory of processing that includes a grammatical theory T and parsing procedure P (say, a procedure that assumes that operations are serial and additive, in that each operation adds a fixed “cost”), then it is T that is challenged and must be changed. The conclusion is particularly unreasonable in the light of the fact that in general there is independent (so-called “linguistic”) evidence in support of T while there is no reason at all to believe that P is true. (Chomsky, 1981: p. 283)

The main methodological implication of this position, which I will refer to as the *strong C/PD*, is that it leads to what I call the ‘*Chomskian paradox*’. On the one hand, the strong C/PD makes T immune to all empirical falsification, since any falsifying evidence can always be dismissed as a

the standard way of implementing symbolic recursion in Von Neumann architectures—i.e., using a push-down stack and multiple subroutines—in their modelling of the human sentence processing mechanisms. This kind of connectionist solution to the problem of recursion in NL behaviour is orthogonal to the subject of this paper, since it merely implies a non-symbolic *implementation* of a symbolic model.

consequence of a false P. However, on the other hand, all grammatical theories rely on grammaticality judgements that (indirectly via processing) display our knowledge of language. Consequently, it seems paradoxical that only certain kinds of empirical material is accepted—i.e., grammaticality judgements—whereas other kinds are dismissed on what appears to be relatively arbitrary grounds. Thus, the strong C/PD provides its proponents with a protective belt that surrounds their grammatical theories and makes them empirically impenetrable to psycholinguistic counterevidence.

In contrast, a more moderate position, which I will refer to as the *weak C/PD*, contends that although linguistic competence is supposed to be infinite, the underlying grammar must support an empirically appropriate performance. This is done by explicitly allowing performance—or processing—considerations to constrain the grammar. Pickering & Chater (1992) have suggested that such constraints must be built into the representations underlying the grammatical theory, forcing a closer relation to the processing theory. This ensures that the relation between the theory of grammatical competence (Chomsky’s T) and the processing assumptions (Chomsky’s P) is no longer arbitrary, resulting in an opening for empirical testing. Nevertheless, inasmuch as T and P are still functionally independent of each other, the option is always open for referring any falsifying empirical data questioning T to problems regarding the independent P, i.e., to performance errors.

To compare the methodological differences between models of NL processing that adopt, respectively, the strong or the weak C/PD, it is illustrative to conceptualize the models as rule-based production systems. In such a system, the grammar would correspond to a knowledge base consisting of a set of declarative rules, each corresponding to a rule in the grammar. The system has a working memory (WM) in which intermediate processing results are stored. The content of the WM is changed by the system through the application of the rules in its knowledge base. A rule can be applied when its right-hand side matches the current content of the WM (or an appropriate part of it).<sup>3</sup> For example, if the content of the WM consists of the two words, say, the<sub>Det</sub> and dog<sub>N</sub> the system would be able to apply a rule such as NP → Det N changing the content of WM to, say, [<sub>NP</sub> the<sub>Det</sub> dog<sub>N</sub>].

Within this framework, the grammar of a particular linguistic theory corresponds to the system’s knowledge base. The system can therefore be said to have an infinite linguistic competence in virtue of its independent knowledge base, whereas its per-

<sup>3</sup>Although it is assumed here that we are dealing with a bottom-up parser, no significant changes would have to be made were we to parse top-down instead.

formance through processing is constrained by WM limitations. This is in direct correspondence with the strong C/PD, since the grammar is completely separated from processing. Models adhering to the weak C/PD would similarly have an independent, declarative knowledge base corresponding of the grammar, but in addition they would also have an extra knowledge base consisting of what we might coin *linguistic meta-knowledge*. This knowledge consists of various performance motivated parsing heuristics that provide context-dependent constraints on the *application* of grammatical rules—such as, for example, the ‘*minimal attachment principle*’ (Frazier & Fodor, 1978). Thus, the performance of the model is constrained not only by limitations on WM but also by linguistic meta-knowledge.

From the production system analogy it can be seen that proponents of both the strong and the weak C/PD stipulate grammars that are functionally independent from processing. As a consequence, empirical evidence that appears to falsify a particular grammar can always be rejected as a result of processing constraints—either construed as limitations on WM (strong C/PD) or as a combination of WM limitations and false linguistic meta-knowledge (weak C/PD). In short, as long as the C/PD—weak or strong—is upheld, potentially falsifying evidence can always be explained away by referring to performance errors. This is methodologically unsound insofar as linguists want to claim that their grammars have representational reality. By evoking the distinction between grammatical competence and observable NL behaviour, thus disallowing negative empirical testing, they cannot hope to find other than speculative support for their theories. In other words, if linguistic theory is to warrant representational claims, then the C/PD will have to be abandoned.<sup>4</sup>

In contrast, a connectionist perspective on NL promises to eschew the C/PD, since it is not possible to isolate a network’s representations from its processing. The relation between the “grammar”, which has been acquired through training, and the processing is as direct as it can be (van Gelder, 1990b). Instead of being a set of *passive* representations of declarative rules waiting to be manipulated by a central executive, a connectionist grammar is distributed over the network’s memory as an *ability to process language* (Port & van Gelder, 1991). In this connection, it is important to notice that although networks are generally “tailored” to fit the linguistic data, this does not simply imply

<sup>4</sup>By this I do not mean that the present linguistic theories are without explanatory value. On the contrary, I am perfectly happy to accept that these theories might warrant certain *indirect* claims with respect to the language mechanism, insofar as they provide means for describing empirical NL behaviour.

that a network’s failure to fit the data is passed on to the processing mechanism alone. Rather, when you tweak a network to fit a particular set of linguistic data, you are not only changing how it will *process* the data, but also what it will be able to *learn*. That is, any architectural modifications will lead to a change in the overall constraints on a network, forcing it to adapt differently to the contingencies inherent in the data and, consequently, to the acquisition of a different grammar. Thus, since the representation of the grammar is an inseparable and *active* part of a network’s processing, it is impossible to separate a connectionist model’s competence from its performance.

However, this leaves open the question of what kind of performance data should be accepted. For the purpose of empirical tests of NL mechanisms we need to distinguish between ‘real’ performance data as exhibited in normal NL behaviour and examples of abnormal or ‘*pathological*’ performance such as ‘slips-of-the-tongue’, blending errors, etc. It might be objected that by proposing such a distinction I am letting the C/PD in by the back door. Yet, this is not the case, since we can plausibly assume that the language processor is an informationally encapsulated, modular system and that pathological performance is due to factors outside the language module.

In this way, what counts as valid data is not dependent on an abstract, idealized notion of linguistic competence but on observable NL behaviour under statistically ‘normal’ circumstances. Consequently, we should be able to filter out the pathological performance data from a language corpora simply by using ‘weak’ statistical methods. For example, Finch & Chater (1992) applied simple bigram statistics to the analysis of a *noisy* corpus consisting of 40,000,000 English words and were able to find phrasal categories defined over similarly derived approximate syntactic categories. It seems very likely that such a method could be extended to a clausal level in order to filter out pathological performance data. Thus, having suggested what qualifies as empirical evidence with respect to models of NL behaviour, I will discuss below whether such data warrant a recursive processing mechanism.

### Recursion and Natural Language Behaviour

The history of the relationship between grammar and language mechanism dates back to Chomsky’s (1957) demonstration that language can, in principle, be characterized by a set of generative rules.<sup>5</sup> In addition, he argued that NL cannot be accounted for by a finite state automaton, because the latter can only produce regular languages. This class of

<sup>5</sup>For a detailed historical overview, see Pickering & Chater (1992).

languages—although able to capture *left-* and *right-*embedded recursive structures—cannot represent *centre-*embedded expressions.<sup>6</sup> For linguistic theories adhering to the C/PD (weak or strong), such a restriction on the power of the finite-state grammars prevents them from being accepted as characterizations of the idealized linguistic competence. On this view, NL must be at least context-free, if not weakly context-sensitive (cf. Horrocks, 1987). However, having eschewed the C/PD, the question is how much processing power is needed in order to account for observable NL behaviour. Do we need to postulate a NL mechanism with the full computational power of a recursive context-free grammar?

Before answering this question, it is worth having a look at some examples of different kinds of recursive NL expressions. Since the crucial distinction between regular and other richer languages is that the former cannot produce expressions involving unbounded centre-embedding, we will look at such sentences first. As the following three examples show, the difficulty of processing a centre-embedded sentence increases with the depth of embedding:

- (1) The boy the girl saw fell.
- (2) The boy the girl the cat bit saw fell.
- (3) The boy the girl the cat the dog chased bit saw fell.

The difficulty of understanding such centre-embedded sentences has been the subject of much debate (e.g., Frazier & Fodor, 1978; Kimball, 1973; Pulman, 1986; Reich, 1969; Wanner, 1980). Proponents of the C/PD have explained the difficulty in terms of performance limitations. For example, in order to account for the problems of parsing recursively centre-embedded sentences, both Kimball's (1973) parser and Frazier's & Fodor's (1978) 'Sausage Machine' parser apply a performance-justified notion of a viewing 'window' (or look-ahead). The window, which signifies memory span, has a length of about six words and is shifted continuously through a sentence. Problems with centre-embedded sentences are due to the parser not being able to attach syntactic structure to the sentences because the verb belonging to the first NP is outside the scope of the window. However, this solution is problematic in itself (cf. Wanner, 1980) since triply centre-embedded sentences with only six words do exist and are just as difficult to understand as longer sentences of similar kind; e.g.,

- (4) Boys girls cats bite see fall.

<sup>6</sup>I will adopt the standard notion of these three kinds of embedded recursion. In case  $X$  is a non-terminal symbol, and  $\alpha$  and  $\beta$  are strings of terminal and non-terminal symbols, we have left-embedding when  $X \Rightarrow X\beta$  (i.e., there is a derivation from  $X$  to  $X\beta$ ), a centre-embedding when  $X \Rightarrow \alpha X\beta$ , and a right-embedding when  $X \Rightarrow \alpha X$ .

A plausible way out of this problem due to Reich (1969) is to argue that centre-embedded sentences, such as (1)–(4), are ungrammatical. Pulman (1986) has opposed this move by contending that with increased computational resources (e.g., pen and paper) or practice, performance on centre-embedded sentences generally increases, whereas this is not the case for ungrammatical strings. However, when we abandon the C/PD, the distinction between 'grammatical' and 'ungrammatical' becomes less important, since we seek to account for performance data as exhibited by typical NL behaviour, rather than abstract grammatical competence. Thus, the difficulty encountered when parsing centre-embedded sentences suggests that NL models need to display the same problems when confronted with this kind of recursive expressions. Still, this solution leaves left- and right-recursion to be dealt with. That these structures cannot be easily dismissed, but seem to be relatively ubiquitous in NL, can be seen from the following examples involving such phenomena as multiple prenominal genitives (5), right-embedded relative clauses (6), multiple embeddings of sentential complements (7), and PP modifications of NPs (8):

- (5) [[[[Bob's] uncle's] mother's] cat]...
- (6) [This is [the cat that ate [the mouse that bit [the dog that barked]]]].
- (7) [Bob thought [that he heard [that Carl said [that Ira was sick]]]].
- (8) ... the house [on [[the hill [with the trees]]][at [the lake [with the ducks]]]].

Furthermore, *prima facie* there seems to be no immediate limits to the length of such sentences.

Even though (5)–(8) are describable in terms of left- or right-recursion, it has been argued—with support from, e.g., intonational evidence (Reich, 1969)—that these expressions are not recursive but *iterative* (Ejerhed 1982; Pulman, 1986). In case these structures are iterative, rather than recursive, then it is possible to account for NL solely in terms of a finite state automaton (FSA). Strong support for this claim comes from Ejerhed (1982) who demonstrated that it is possible for a FSA, comprising a non-recursive context-free grammar, to capture the empirical data from Swedish (provided that unbounded dependencies are dealt with semantically). This demonstration is significant because Swedish is normally assumed to require the power of context-sensitive languages (e.g., cf. Horrocks, 1987). Thus, we have strong reasons for believing that a non-recursive FSA provides sufficient computational power to account for NL performance without needing to postulate a functionally independent infinite competence. As we shall see below, certain kinds of connectionist models, that is, simple recurrent networks, have the ability to mimic FSAs in a psychologically interesting way.

## A Connectionist Account of “Recursive” Natural Language Behaviour

We have found that it is not necessary to invoke the C/PD in order to account for NL processing. Moreover, we have seen that a non-recursive FSA has sufficient computational power to function as a NL processing mechanism. So, the remaining question is whether a connectionist model can mobilize such power—or whether we have to give in to Fodor’s & Pylyshyn’s (1988) negative claims concerning connectionist NL processing. In the rest of this paper, I provide arguments to the effect that a particular kind of connectionist model—the simple recurrent network (SRN) (Elman, 1990, 1991)—promises to have sufficient power to capture NL behaviour.

An SRN is a connectionist feed-forward network that has an extra set of hidden, so-called ‘*context*’ units (Elman, 1990, 1991). At time  $t$ , the hidden unit activation is copied over into the context units. Via recurrent links, the activation over the context units is fed back (as part of the input) to the hidden units at time  $t + 1$ . In this way, the presence of recurrent links, together with the context units, allows past activation to influence the current output, thus enabling the network to encode temporal sequences. The latter is typically encoded in terms of a prediction task in which the SRN is trained to predict the next item in a sequence (e.g., the next word in a sentence).

Simulation results obtained by Servan-Schreiber, Cleeremans & McClelland (1991) show that an SRN is able to mimic an FSA in a quite unique way. Instead of encoding the discrete finite representations corresponding to particular inputs, as in a traditional FSA, the network encodes an *association* between a given input and the appropriate prediction of the next output state. This allows the network to capture long-distance dependencies by *shading* its internal representations; that is, by picking up subtle statistical contingencies. In other words, the network learns to respond to temporally distant information by encoding contextually relevant cues in a condensed form in the recurrent links. In addition, SRNs appear to have *functional compositionality* (van Gelder, 1990a) insofar as they are able to process functionally compound representations in a way that is sensitive to their constituent structure. These results demonstrate that SRNs have sufficient power to develop representations that possess the rich internal structure that is necessary for the explanation of systematic NL behaviour.

In a particularly interesting simulation, Elman (1991) demonstrated that a SRN, though inherently sensitive to context, can learn the abstract and general grammatical structure implicit in a language corpus. The network was trained on four

different corpora, each of which had an increasing number of relative clauses, and which together totalled 10,000 sentences with a length between 2 and 16 words. The distributed representations developed by the network through training were analyzed in terms of the trajectories through state space over time. More specifically, the trajectories correspond to the internal representations evoked at the hidden unit layer, as the network processed a given sentence (Elman, 1991). Elman’s analysis showed that the network was able to capture *agreement* between subject nouns and verbs. The network also developed *verb argument structure*; that is, the network learned to behave in an appropriate manner according to whether it encountered intransitive, transitive, or optionally transitive verbs.

From the viewpoint of the present paper, the most interesting result of this simulation was that the network developed a differentiated capacity with respect to the processing of complex sentences with *recursive* structure. For example, the network was able to process the following *centre-embedded* sentence involving long-distance agreement dependencies:

(9) Boys who girls who dogs chase see hear.

Trajectory-analysis of similar sentences evinced that successive embedded clauses are represented in the same way as the first embedded clause, but slightly displaced in state space. This systematic displacement of recursive clauses in state space enabled the network to keep track of the depth of recursion, while at the same time acknowledging structural similarities between the recursive clauses. However, the network’s performance on recursive sentences was limited. An interesting fact about the network’s degrading recursive performance was that sentences involving centre-embedded recursion were more badly affected than sentences involving right-embedding. This is consonant with our earlier psycholinguistic observations regarding the parsing of recursive structures.

Pace Fodor & Pylyshyn (1988), connectionist models are suitable for the modelling of NL processing. Indeed, as we have seen, the SRN is particularly interesting from a psycholinguistic perspective in that it appears to exhibit the same behaviour as humans when confronted with complex, recursive sentences (Elman, 1991)—without reverting to *explicitly* programmed limitations on memory. The simulations conducted by Servan-Schreiber, Cleeremans & McClelland (1991) have shown that an SRN is able to mimic a graded FSA—but, more importantly, the SRN *learns* how to behave as if it was an FSA *with a limited stack*, enabling it to deal with centre-embedded sentences with a limited depth of nesting. In contrast, performance oriented symbolic approaches to NL processing (typically also based on FSAs) need to build in such

limitations *explicitly*; for example, as limitations on the number of iterations in a regular grammar (Reich, 1969), or as a procedure that clears a stack-like memory structure when a certain threshold is met (Pulman, 1986). Hence, connectionist models provide an appealing non-symbolic account of recursion in linguistic descriptions, while respecting actual psycholinguistic constraints on human NL processing. Crucially, performance aspects do not have to be programmed explicitly outside a connectionist model—they, so to speak, “fall” out in a natural way as a side-effect of the processing of recursive sentences.

### Conclusion

In this paper, I have argued that recursion in NL is best construed as a descriptive phenomenon, rather than a basic processing mechanism. In addition, I have questioned the psychological plausibility of the C/PD and have advocated the incorporation of psychological constraints into the NL processing mechanism. It is possible for a connectionist model to account for recursion in NL insofar as the notion of infinite competence is dropped and replaced with a psychologically constrained processing ability. However, such a model must be able to explain empirical data from NL behaviour as an interaction between processing abilities and limitations inherent in the model itself. Recursion in NL is therefore only a problem insofar as linguistic theories are viewed as having explanatory adequacy, and insofar as the notion of an infinite competence is maintained. Work within the connectionist paradigm indicates that descriptive recursion can be accounted for in a way which follows empirical constraints on NL behaviour. However, it is too early to say whether connectionism in the long run will be able to account for the full complexity of human NL behaviour. However, at least presently, connectionism provides a promising framework for non-symbolic NL research.

### Acknowledgements

Many thanks to the members of the Foundations of Cognitive Science workshop at the Centre for Cognitive Science, especially Nick Chater and Martin Pickering, for comments and suggestions regarding earlier drafts of this paper. Thanks are also due to Elisabet Engdahl and Ewan Klein for commenting on the penultimate draft.

### References

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Forris Publications.  
 Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger  
 Ejerhed, E. (1982). The Processing of Unbounded Dependencies in Swedish. In E. Engdahl & E.

Ejerhed (eds.), *Readings on Unbounded Dependencies in Scandinavian Languages*. Stockholm: Almqvist & Wiksell International.  
 Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, **14**, 179-211.  
 Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, **7**, 195-225.  
 Finch, S. & Chater, N. (1992). Bootstrapping Syntactic Categories by Unsupervised Learning. Forthcoming in Proceedings of the 14th Annual Conference of the Cognitive Science Society.  
 Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, **28**, 3-71.  
 Frazier, L. & Fodor, J.D. (1978). The Sausage Machine: A New Two Stage Parsing Model. *Cognition*, **6**, 291-325.  
 Gazdar, G., Klein, E., Pullum, G. & Sag, I. (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell.  
 Horrocks, G. (1987). *Generative Grammar*. London: Longman.  
 Kimball, J. (1973). Seven Principles of Surface Structure Parsing in Natural Language. *Cognition*, **2**, 15-47.  
 McClelland, J. & Kawamoto, A.H. (1986). Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. Chapter 19 in J. McClelland & D. Rumelhart (eds.), *Parallel Distributed Processing*, Volume 2. Cambridge, Mass.: MIT Press.  
 Pickering, M. & Chater, N. (1992). Processing Constraints on Grammar. Ms.  
 Port, R. & van Gelder, T. (1991). Representing Aspects of Language. In Proceedings of the 13th Meeting of the Cognitive Science Society, 487-492, Chicago, Illinois: Cognitive Science Society.  
 Pulman, S.G. (1986). Grammars, Parsers, and Memory Limitations. *Language and Cognitive Processes*, **2**, 197-225.  
 Reich, P. (1969). The Finiteness of Natural Language. *Language*, **45**, 831-843.  
 Servan-Schreiber, D., Cleeremans, A. & McClelland, J. (1991). Graded State Machines: The Representation of Temporal Contingencies in Simple Recurrent Networks. *Machine Learning*, **7**, 161-193.  
 van Gelder, T. (1990a). Compositionality: A Connectionist Variation on a Classical Theme. *Cognitive Science*, **14**, 355-384.  
 van Gelder, T. (1990b). Connectionism and Language Processing. In G. Dorffner (ed.) *Konnectionismus in Artificial Intelligence und Kognitionforschung*. Berlin: Springer-Verlag.  
 Wanner, E. (1980). The ATN and the Sausage Machine: Which One is Baloney? *Cognition*, **8**, 209-225.