

LSA Annual Meeting
Atlanta, Georgia
January 3, 2003

Connectionism and Dynamics
in the special session on Probability Theory in Linguistics
Whitney Tabor
Department of Psychology, University of Connecticut
tabor@uconnvm.uconn.edu
<http://www.sp.uconn.edu/ps300vc/tabor.html>

Introduction

Two research projects:

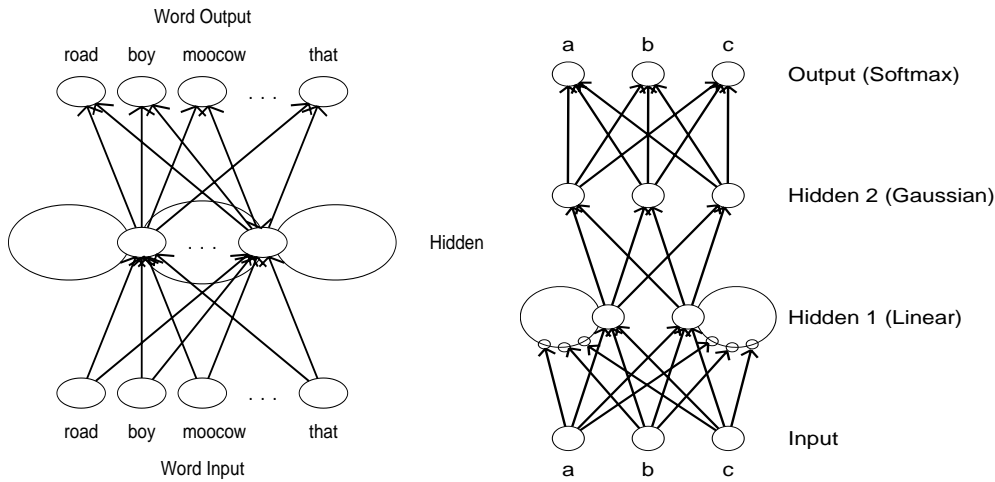
1. Learning complex phrase-structure grammar with connectionist networks (linguistics)
2. Modeling sentence-processing with self-organizing constituent attachment networks (psycholinguistics)

Common themes: **recurrent connectionist networks, attractor landscapes, self-organization**

Conclusion: it is not so much probabilities as activations that fill in the "continuum between the absolutes"

Project 1. Phrase structure grammar learning (linguistics)

- (1) ...this moocow that was coming down along the road met a nicens little boy named baby tuckoo... (Joyce, 1917)
 - (2)
 - IP \rightarrow NP VP
 - NP \rightarrow Det N'
 - N' \rightarrow N CP
 - CP \rightarrow Comp IP
 - (3) Motivations:
 - (1) The phrase structure base is fundamental to all current syntactic theories and yet there is no effective formal theory of how we know what it is.
 - (2) Many of the most challenging and interesting problems in linguistics are constituency problems.
 - (4) Simple Recurrent Networks (SRNs) and variants on them can represent and even *learn* recursive structures (Wiles & Elman, 1995; Rodriguez, Wiles, & Elman, 1999; Bodén & Wiles, 2000, 2002; Gers & Schmidhuber, 2001)
- Pace Pinker (1999), Steedman (1999), Marcus (2001), Jackendoff (2002) who have suggested that these kinds of networks are incapable of learning recursive structures.

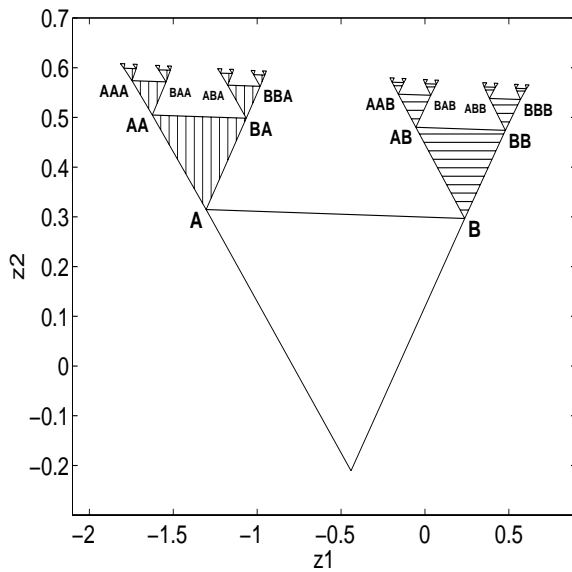


(a) Simple Recurrent Net (SRN) (b) Fractal Learning Neural Net (FLNN)

Language	Sample sentences
$a^n b^n$	a b, a a b b, a a a b b b
$a^n b^n c^n$	a b c, a a b b c c, a a a b b b c c c
$a^n A^m B^m a^n$	a b, a A B b, a a A B b b, a A A B B b

- (5)
- (6) Learning exponential state-growth languages is much more difficult (Elman, 1991, 1993; Rohde & Plaut, 1999; Rodriguez, 2001)
- 1.0 $S \rightarrow A B C$
- 1.0 $A \rightarrow a (S)$
- 1.0 $B \rightarrow b (S)$
- 1.0 $C \rightarrow c (S)$

(7) Visited states in the hidden unit space of the FLNN:



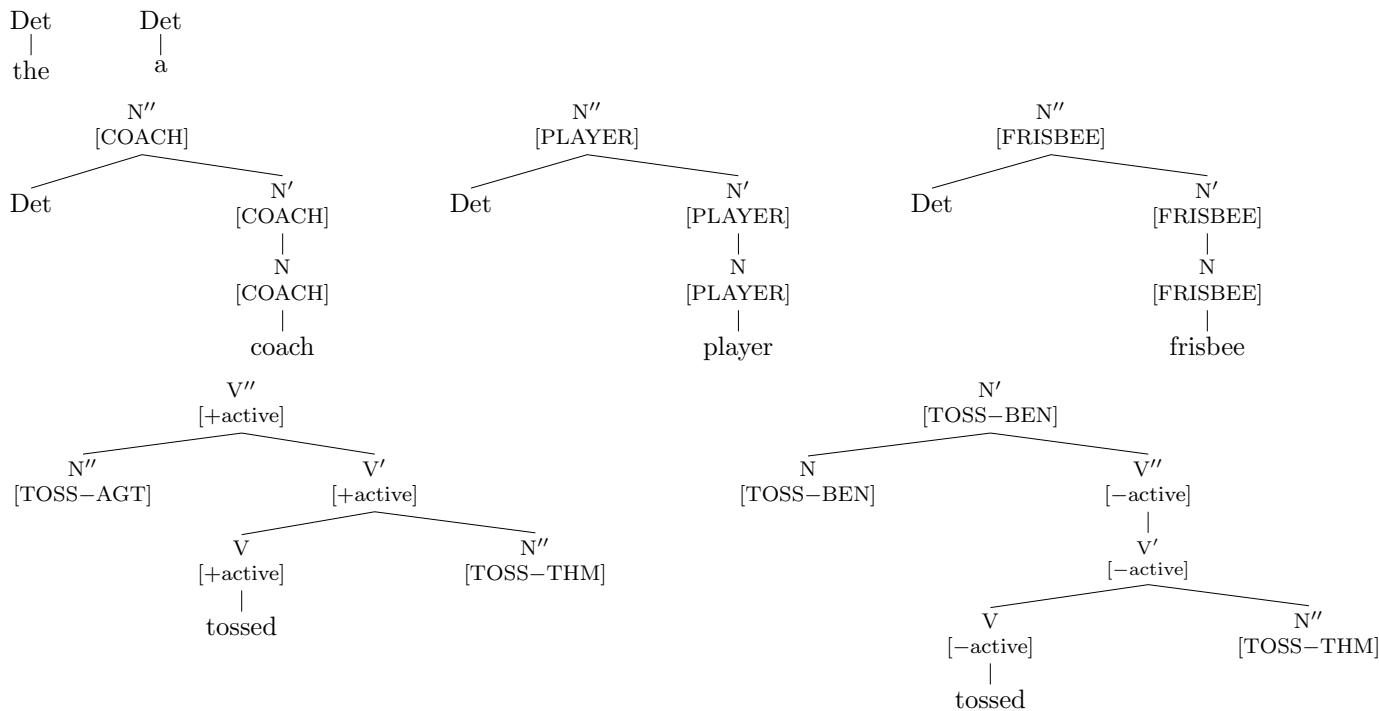
(8) How the FLNN works:

1. Attactors exert "pulses" of influence on the system state, coloring it with its history.
2. Fractal set organizes computation: analogous to recursive rules.
3. Self-organizing in the sense that the interactions of autonomous lexical entities give rise to the syntactic structure.
4. The learning is a kind of tuning of probabilities, but every finite word-sequence has positive probability (so the learning is not equivalent to tuning the probabilities on a restrictive phrase-structure grammar).

Project 2: Sentence processing (psycholinguistics)

(9) Self-organized parsing: Kempen & Vosse (1989), Stevenson (1994, 1997a, 1997b, 1998), Vosse & Kempen (2000):

(10) Self Organizing Tree Fragments:

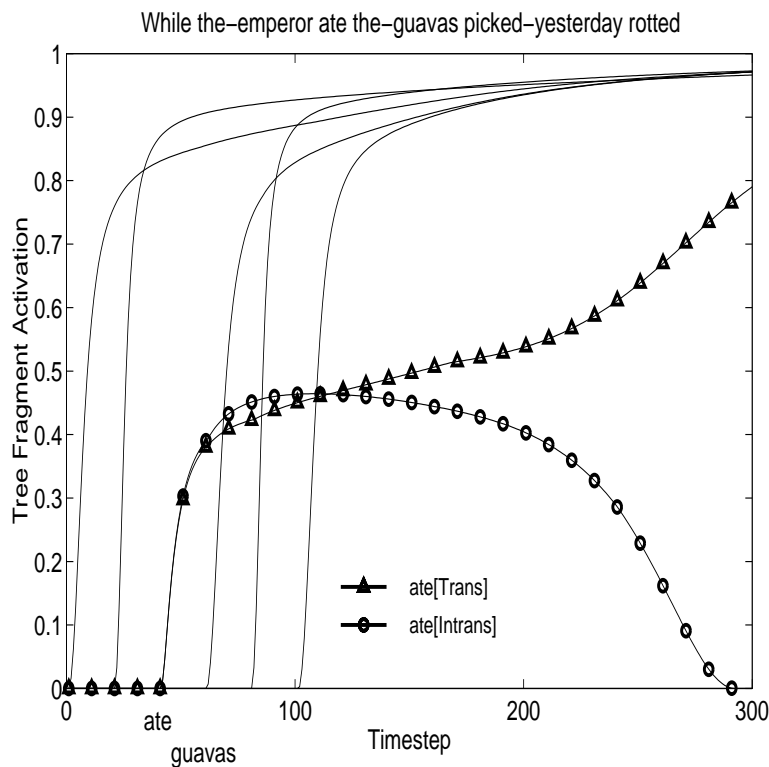


(11) How the self-organizing parser works:

1. Links grow at rates proportional to the feature match between their nodes.

2. Multiple senses (syntactic and semantic) of ambiguous words compete.
3. The attractors are the stable states that the system "gravitates toward"
4. Reading time is stabilization time.
5. The system is, again, self-organizing: syntactic structures are built up out of a collection of lexical entities vying to form combinations with one another (there is no global well-formedness requirement)
6. Lexical biases are implemented as initial activation differences.

(12) Processing a simple garden path sentence:



2.1 Coverage of well-known sentence-processing results

- (13) Center-embedding difficulty (Vosse & Kempen, 2000)
 - a. The cat the rat chased died. (100% successfully parsed)
 - b. The cat the rat the dog bit chased died. (0% successfully parsed)
- (14) **Reason:** the number of distracting potential attachments grows exponentially with the number of embeddings.

(15) Lexical biases (McRae, Spivey-Knowlton, & Tanenhaus, 1998):

- a. The cop arrested by the detective was chagrined.
 FAST SLOW
- b. The crook arrested by the detective was chagrined.
 SLOW FAST

(16) **Reason:** Felicitous semantic combinations favor some links over others.

(17) Complexity effects (Grodner, Gibson, & Tunstall, 2002):

The witness who the evidence examined...

- a. ...by the prosecutor offended made a complaint. (HARD)
- b. ...made a complaint. (EASIER)

(18) **Reason:** Complex trees are outcompeted by simple trees.

2.2 Distinguishing predictions of the self-organizing model.

(19) Lexical biases interact with center-embedding (King & Just, 1991; Weckerly, 1995; Traxler, Morris, & Seeley, 2002):

- a. The doorman that the chauffeur greeted called to the dog. (HARD at "called")
- b. The limousine that the chauffeur drove needed extensive work. (EASY at "needed")

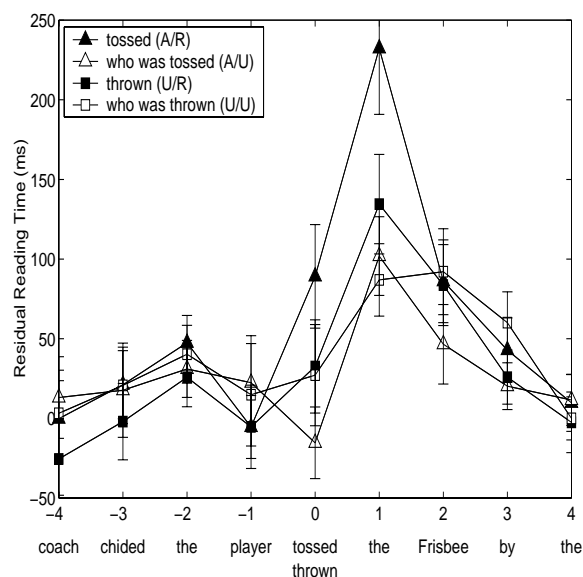
(20) **Self-organizing model:** semantic biases support appropriate link-growth in the subordinate clause, easing the stress at the main verb.

Statistical models based on grammatical parses: no difference.

(21) Distracting local coherence (Tabor & Galantucci, 2000; Tabor, 2002b):

- a. The coach chided the player tossed the frisbee by the opposing team.
(AMBIGUOUS, REDUCED)
- b. The coach chided the player who was tossed the frisbee by the opposing team.
(AMBIGUOUS, UNREDUCED)
- c. The coach chided the player thrown the frisbee by the opposing team.
(UNAMBIGUOUS, REDUCED)
- d. The coach chided the player who was thrown the frisbee by the opposing team.
(UNAMBIGUOUS, UNREDUCED)

(22) Interaction between Ambiguity and Reduction indicates formation of local agentive structure corresponding to "the player tossed the frisbee" in (21a):



- (23) **Self-organizing model:** because of self-organization, locally coherent combinations form and temporarily compete with the globally optimal parse.

Statistical models based on grammatical parses: Reduction effect due to constraining information of "who was". Possible ambiguity effect. No interaction.

- (24) Length effects (Warner & Glass, 1987; Ferreira & Henderson, 1991, 1993, 1998; Tabor, 2002a):

- a. While the boy scratched the dog yawned. (EASIER)
- b. While the boy scratched the dog sleeping peacefully yawned. (HARDER)

- (25) Self-organizing model: length effect due to decay (Stevenson, 1998) or bond entrenchment (as in the self-organizing model outlined above).

Statistical models based on grammatical parses: No difference.

3. Conclusions

- (26) The language processor is a lexically-based, self-organizing system.
- (27) A major advantage of the self-organizing approach is that it allows a number of processing results (hitherto handled by diverse processing-specific mechanisms) to stem from interaction between linguistic structures and general dynamical constraints.
- (28) Activations, which are approximated by, but are not equivalent to probabilities fill in the "continuum between the absolutes".

References

- Bodén, M., & Wiles, J. (2000). Context-free and context sensitive dynamics in recurrent neural networks. *Connection Science*, 12(3), 197–210.
- Bodén, M., & Wiles, J. (2002). On learning context-free and context-sensitive languages. *IEEE Transactions on Neural Networks*, 13(2), 491–493.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71–99.
- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725–745.
- Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology*, 47(2), 247–275.
- Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73–100). Dordrecht: Kluwer Academic Publishers.
- Gers, F. A., & Schmidhuber, J. (2001). Lstm recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks*, 12(6), 1333–1340.
- Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, 46, 267–295.
- Jackendoff, R. (2002). *Precis of: Foundations of language: Brain, meaning, grammar, evolution*. (To appear in *Behavioral and Brain Sciences*)
- Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulated annealing. *Connection Science*, 1, 273–290.
- King, J., & Just, M. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Marcus, G. F. (2001). *The algebraic mind*. Cambridge, MA: MIT Press.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in online sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Pinker, S. (1999). *Words and rules: the ingredients of language*. New York: Perennial.
- Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation*, 13(9).

- Rodriguez, P., Wiles, J., & Elman, J. (1999). A recurrent neural network that learns to count. *Connection Science*, *11*(1), 5–40.
- Rohde, D., & Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Journal of Memory and Language*, *72*, 67–109.
- Steedman, M. (1999). Connectionist sentence processing in perspective. *Cognitive Science*, *23*(4), 615–636.
- Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, *23*(4), 295–322.
- Stevenson, S. (1997a). Lexical structure and parsing complexity. *Language and Cognitive Processes*, *12*(2/3), 349–399.
- Stevenson, S. (1997b). An intergrated symbolic/connectionist parsing archicture. In R. Sun & F. Alexandre (Eds.), *Connectionist-symbolic integration: From unified to hybrid approaches* (pp. 209–206). Mahwah, NJ: Lawrence Erlbaum Assoc.
- Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 327–363). Dordrecht: Kluwer Academic Publishers.
- Tabor, W. (2002a). *Evidence for self-organized sentence processing: Digging-in effects*. (Submitted manuscript: See <http://www.sp.uconn.edu/ps300vc/papers.html>)
- Tabor, W. (2002b). *Self-organized sentence processing*. (Submitted manuscript: See <http://www.sp.uconn.edu/ps300vc/papers.html>)
- Tabor, W., & Galantucci, B. (2000). Ungrammatical influences: Evidence for dynamical language processing. In *Proceedings of the 22nd annual meeting of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Traxler, M. J., Morris, R. K., & Seeley, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*, 69–90.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.
- Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*, 714–738.
- Weckerly, J. (1995). *Object relatives viewed through behavioral, electrophysiological, and modeling techniques*. (Ph.D. dissertation, University of California, San Diego)
- Wiles, J., & Elman, J. (1995). Landscapes in recurrent networks. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual cognitive science conference*. Lawrence Erlbaum Associates.