A Dynamical System for Language Processing

Whitney Tabor[1,2] and **Cornell Juliano**[1] and **Michael Tanenhaus**[1]
[1]Department of Brain and Cognitive Sciences
[2]Center for the Sciences of Language
University of Rochester
Rochester, NY 14627
whitney@bcs.rochester.edu

Abstract

A dynamical systems model of language processing suggests a resolution of the debate about the influences of syntactic and lexical constraints on processing. Syntactic hypotheses are modeled as *attractors* which compete for the processor's trajectory. When accumulating evidence puts the processor close to an attractor, processing is quick and lexical differences are hard to detect. When the processor lands between several attractors, multiple hypotheses compete and lexical differences can tip the balance one way or the other. This approach allows us to be more explicit about the *emergent properties* of lexicalist models that are hypothesized to account for syntactic effects (MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell & Tanenhaus, 1994).

Introduction

Readers and listeners have clear preferences for certain syntactic sequences (e.g., NVN as a main clause), as revealed by garden-path effects for temporarily ambiguous sentences that do not conform to these preferences (1).

(1) a. The horse raced past the barn fell.
b. The patient warned the doctor was incompetent.

However, recent evidence suggests that these garden-path effects can be sharply reduced by strong lexical constraints, as illustrated by the examples in (2), which are structurally similar to those in (1), but do not appear to cause processing difficulty.

(2) a. The land mine buried in the sand exploded.
b. The patient said the doctor was incompetent.

The interpretation of these lexical effects has been extremely controversial. In the influential class of "structure-first" models, category-based parsing phenomena are accounted for by positing an initial, encapsulated processing stage in which structure is built using syntactic category information and a few, general principles (e.g., Frazier, 1987). Lexically-specific information and other (non-syntactic) constraints apply at a later stage in processing. In contrast, several research groups have argued that many of the phenomena that motivated these category-based principles can be reduced to the effects of interacting lexical constraints (McClelland, St. John, & Taraban, 1989; MacDonald, Pearlmutter, & Seidenberg 1994; Trueswell & Tanenhaus 1994).

The competition between the lexicalist and structuralist claims has focused increased attention on a long-standing, empirical debate about the time-course with which lexical constraints are observed relative to structural constraints during on-line sentence processing (for a recent review see Tanenhaus & Trueswell, 1995). While the results of these experiments have often been equivocal, two generalizations emerge. First, the relative strength of structural and lexical constraints varies across contexts and structures. Second, there are clear circumstances in which local lexical constraints are insufficient to capture important processing generalizations (e.g, island constraints on movement and structural preferences even in the face of strong contrary lexical biases).

Research on automatic parsing has led to a class of models which use corpus-tuned probabilistic grammars to compute *conditional probabilities* of lexical items in contexts (see Charniak, 1993, for review). Recently, models of this type have been used to combine lexical and syntactic information to make reading time predictions (Jurafsky, 1996). These models provide a theoretical basis for incorporating probabilistic lexical information in a model that uses syntactic rules, but they do not provide insight into the variation across contexts of the relative strengths of structural and lexical constraints.

A promising approach to explaining this variation is to treat category-based parsing preferences as generalizations that emerge within a constraint-based learning system because of similarities among "classes" of lexical items (e.g., Juliano & Tanenhaus, 1994). But prior proposals along these lines have been vague. In this paper we show how certain constructs of dynamical systems theory allow us to be more explicit about the nature of the "emergent" representations, their relationship to traditional syntactic categories, and their empirical predictions.

Dynamical Systems Theory

Dynamical systems theory (see Abraham and Shaw 1984 and Strogatz 1994 for introductions) is typically concerned with systems that change continuously with time. Examples of much-studied dynamical systems include: pendulums swinging on rigid arms; stars and planets orbiting one another in space; populations fluctuating in an ecosystem; gases swirling around in the atmosphere. It is useful to consider the trajectories of a dynamical system—i.e., the paths it can follow as time progresses. In the case of a pendulum, some trajectories swing back and forth, others whirl around the circle, and two of them remain at one point indefinitely (hanging down, and, improbably, balanced straight up). If the pendulum is damped, then all trajectories except those leading to the improbable state approach the low point-trajectory in the limit. Such a limiting trajectory is called an *attractor*. Those starting points from which the system gravitates toward a particular attractor A are collectively referred to as the *basin of attraction* of A. The basin of attraction of the pendulum's low attractor consists of every state except those that lead to the improbable state. In a planetary system, each large mass is surrounded by its own basin of attraction. One particularly interesting property of multiple-attractor systems is that when the system is near an attractor A it is dominated by the properties of A alone, but when it is further away, it may still be in the basin of A, but other attractors can exert an influence on it. Below, we use this *local dominance* property of attractors to model the variable balance of syntactic and lexical influences on processing.

Sentence Processing Data

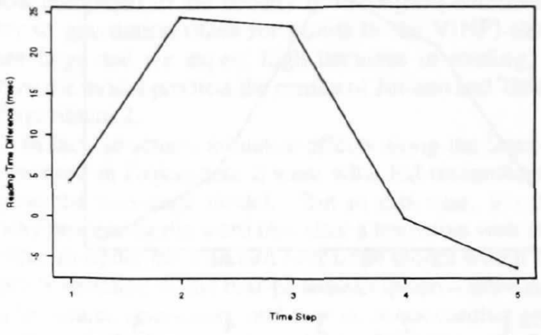
Some recent experiments by Juliano and Tanenhaus (1993) elucidate the way these dynamical systems notions can help simplify our understanding of the relationship of syntactic and lexical influences on processing. Juliano and Tanenhaus focused on the relationship between verbs that take sentence complements (V[Sbar]'s) and verbs that take noun phrase complements (V[NP]'s or *transitive verbs*). Typically, sentence complements are introduced by the complementizer *that*, although the complementizer can be absent (3).

(3) The grocer insisted/agreed/complained/argued (that) the cheap hotel was pleasant.

When the complementizer is absent and the embedded sentence starts with its subject noun phrase, the beginning of the main sentence has the form NP-V-NP (or "NVN" for short), which makes it abstractly consistent with the transitive pattern. It is well known that, in many ambiguous cases, people prefer to interpret a noun phrase after a verb as a direct object (Frazier, 1987). This preference is part of the evidence that motivates the two-stage model of processing: the preference *could* arise because the processor initially assumes that any NP after a verb is the verb's object. Indeed Juliano and Tanenhaus found high latencies at determiners that immediately followed V[Sbar]'s, in comparison to determiners that immediately followed V[NP]'s (Figure 1; see their "Experiment 3" for details). But the effect was correlated with the context-independent frequency of the verb: verbs low in absolute frequency showed a stronger effect than verbs high in absolute frequency (see also Trueswell, Tanenhaus, and Kello, 1993). The frequency correlation is hard to understand under the two-stage model because that model implies that lexical differences have no effect on initial processing. The results suggest a model in which incorrect hypotheses can exert a marginal influence on correct hypotheses to varying degrees, depending on the frequencies of the items involved. We are thus led to the idea of letting syntactic hypotheses correspond to attractors in a dynamical system.

Another experiment by Juliano and Tanenhaus (1993) (their "Experiment 2"), provides further support for the competing attractors approach. Although the transitive pattern is indeed the most common pattern in English declarative-sentence syntax (if we take the Brown Corpus to be representative), it happens to be the case that when an arbitrary verb is followed by the word *that*, the word *that* is most likely to be a complementizer. Therefore, we might hypothesize that the NVN schema has to compete with another schema of the form V-*that*[Comp]. Following the same line of reasoning as before, we thus expect people to have difficulty when the direct object of a transitive verb is introduced by the determiner *that*. Juliano and Tanenhaus (1993) find a strong effect in support of this interpretation: when the same transitive verb is followed by *that* and *the*, reading times at *that* are substantially slower. The effect peaks at the adjective following the determiner and diminishes again when the noun is reached. One might be concerned that this result is due either to the high frequency of *the* (there is generally an inverse correlation between the absolute frequency of a word and its processing time), or to the pragmatic strangeness of using determiner *that* without prior mention of its referent (the sentences were presented without a supporting context). However, the same effect is observed when *that* is compared with *those* (Figure 2). *Those* is less frequent than *that* as a determiner (1:3 in the Brown Corpus) and has similar presuppositions to *that*. Thus, the results still seem best explained by positing influence by the V-*that*[Comp] schema on NVN sentences.

Figure 1: Reading time differences between V[Sbar]-*the* and V[NP]-*the*. (This figure is based on data reported in Juliano and Tanenhaus, 1993, Experiment 3)



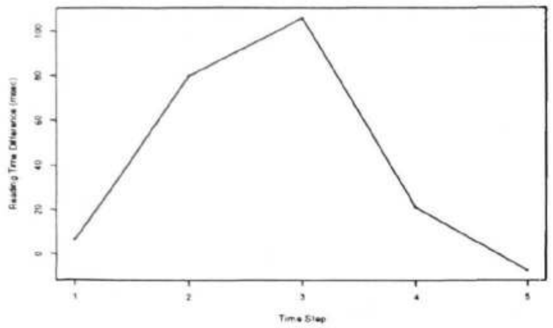
	1	2	3	4	5
	Verb	Det	Adj	Noun	Auxiliary
(a)	insisted	the	cheap	hotel	was. . .
(b)	visited	the	cheap	hotel	to. . .

Both of these results point in the direction of a dynamical systems treatment. In particular, attractors seem like useful devices for modeling the way in which the correct pattern usually "captures" the processor in the long run while incorrect patterns tend influence its behavior along the way. In the next section we develop a computational model which allows us to explicitly model reading times on this basis.

A Dynamical System for Language Processing

We use a vector space to encode the states of the language processor. Each time the processor encounters a word, it jumps to some location in the vector space. We recapitulate most of the distinctions that a symbolic grammar makes by letting isolated regions correspond to symbolic states. Every time the symbolic processor would be in a state S_i, the vector space processor is in a region, R_i. Additionally, we require that the distances between regions reflect partial similarity properties of the data. For example, since PP-complement

Figure 2: Reading time differences between V[NP]-*that* and V[NP]-*those*



	1	2	3	4	5
	Verb	Det	Adj	Noun	Infinitive
(a)	visited	that	cheap	hotel	to. . .
(b)	visited	those	cheap	hotels	to. . .

verbs are relatively similar to transitive verbs in comparison to nouns and prepositions, the region corresponding to the processing of a PP-complement verb is placed relatively near the region corresponding to the processing of a transitive verb. Moreover, within each of the isolated regions, subtle statistical differences between elements are encoded as small within-region contrasts. We refer to the collection of points visited in this representation space during processing of a large sample of language use as a *visitation set*.

Having established the representation space and the visitation set, we assume that processing a word in a corpus involves moving to the appropriate place in the representation space (more on how to do this below) and then migrating to the nearest big cluster, essentially as in a gravitational system. Successful parsing corresponds to arriving at (or getting sufficiently near) a cluster locus. Thus the proposed representation space contains the trajectories of a dynamical system, where the category clusters correspond to attractor basins. Reading time is modeled as the time it takes the processor to gravitate to an attractor. The processor gravitates quickly when it lands near a dense cluster; it gravitates more slowly when it lands near a sparse cluster. Moreover, it reaches the relevant attractor more quickly if it starts near the center of a cluster than if it starts somewhere on the periphery. A verb like, *roll*, positioned on the periphery of the verb cluster because of the competing noun interpretation will take longer to process than an unambiguous verb. Thus the model predicts the general finding that ambiguous elements slow the processor down.

We used a connectionist network for corpus learning to generate the visitation set on which the dynamical model is based. The network had feedforward connections from an input layer through a hidden layer to an output layer and the hidden units were recurrently connected to themselves and one-another (Figure 3). Words were assigned localist representations on the input and output layers and the network was trained using a corpus: each word was presented, in sequence, on the input layer and the weights were adjusted to improve the network's prediction of the successor word (as in Elman, 1991). Error was propagated using backpropagation through time and the error gradient was approximated by attending to only a few, recent timesteps (see Pearlmutter, 1995's review). The hidden unit space of such a network corresponds to the representation space described in the previous paragraph. The visitation set was thus created by collecting a sample of hidden units visited by the network during the processing of a corpus approximating natural usage.

Figure 3: A recurrent network for word prediction. (The activation of layer HIDDEN_{t-1} was set equal to the activation that layer HIDDEN_t had on the previous time-step. The error signal was backpropagated through 3 time-steps.)

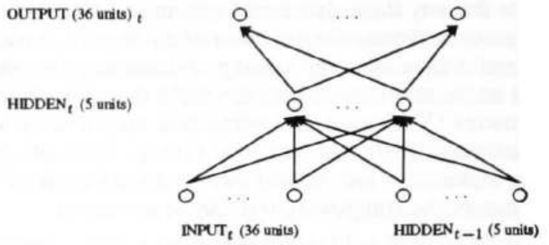


Figure 4: The generating grammar for simulations of Juliano and Tanenhaus's experiments.

1.00	Sroot : S p
1.00	S : NP VP
0.67	VP : VP[NP]
0.33	VP : VP[Sbar]
0.67	VP[NP] : V[NP] NP
0.33	VP[NP] : V[NP]
0.67	VP[Sbar] : V[Sbar] that S
0.33	VP[Sbar] : V[Sbar] S
1.00	NP : Det N
[Zipf]	Det : 0.44 the, 0.22 a, 0.14 which, 0.10 that 0.10 those
[Zipf]	V[NP] : 0.34 called, 0.17 followed, 0.11 pulled, 0.09 caught, 0.07 pushed, 0.06 loved, 0.05 visited, 0.04 studied, 0.04 tossed, 0.03 grabbed
[Zipf]	V[Sbar] : 0.34 thought, 0.17 agreed, 0.11 insisted, 0.09 wished, 0.07 hoped, 0.06 remarked, 0.05 pleaded, 0.04 speculated, 0.04 doubted, 0.03 hinted
[Zipf]	N : 0.34 woman, 0.17 man, 0.11 dog, 0.09 cat, 0.07 blouse, 0.06 hat, 0.05 cake, 0.04 ball, 0.04 watch, 0.03 cypress

To generate a corpus, we used the probabilistic, context free grammar shown in Figure 4. The frequencies of rules in the grammar are set according to a simple rubric called Zipf's Law. Zipf's Law holds that a rank vs. frequency plot of the vocabulary elements drawn from any large corpus forms the cusp of a hyperbola (Zipf, 1943). The law has been confirmed by Zipf and his successors as a fair approximation for numerous corpora in a wide range of languages. We have observed that it also provides a reasonable approximation for several of the major lexical categories in the Brown Corpus (Noun, Verb, Adjective, Adverb, Determiner) and have thus used it to assign frequencies to lexical items in the grammar. Somewhat arbitarily, we have also used it to determine the relative frequencies for the grammar's multi-production syntactic nodes.

After training the network on grammar-generated data until it had learned all of the major distinctions made by the grammar, we formed a visitation set in the following manner. With learning turned off, we let the grammar generate 1000 words in sequence. We presented these to the network in order and formed a visitation set by recording the set of hidden unit locations that the network visited. This visitation set defines the behavior of the dynamical processor. There are a number of slightly different ways of implementing the dynamics. We have experimented with one in which we let each point in the visitation set behave as a point mass and we model the processor as a small test mass which follows a trajectory defined by Newton's Law of Universal Gravitation. The predictions of this model are accurate in the cases we have tried, but it is computationally very expensive. Alternatively, we could let a recurrent neural network form activation-space attractors corresponding to grammatical classifications and treat processing time as relaxation time. Or we could interpret the inverse of density as a potential surface (so that density maxima correspond to valley-bottoms and density minima to hilltops) and model the processor as a ball rolling down this surface. Since we are not yet sure which implementation is best, we take a shortcut here and use an easily computable approximation of attraction time which is most similar to the density-as-potential model but shares the main properties of all the systems mentioned. In particular, for a processing state associated with a juncture, p, between words in a corpus, we let the multiplicative inverse of density provide an estimate of attraction time:

(4) **Def.**

$$Predicted\ Reading\ Time\ at\ p = \frac{1}{1 + Density\ at\ p}$$

For the network trained on the grammar in Figure 4, a two-dimensional projection of part of the visitation set is shown in Figure 5. This projection focuses on cases where a determiner or complementizer occurred after a verb. It was obtained by selecting 1000 cases of determiners/complementizers occurring after verbs and plotting the first two principal components (Jolliffe, 1986) of the corresponding hidden unit locations. The first Principal Component accounts for 97% of the variance, the second for 2%. Thus, the data of Figure 5 are primarily spread out in the x-direction. We show the expanded y-direction for visual clarity.

In this figure, there are two densely populated clusters corresponding to V[Sbar]'s followed by *that* and V[NP]'s followed by an unambiguous determiner. The centers of these two dense clusters are attractors in the dynamical model and they correspond to the two main syntactic patterns involved here: the sentence-complement pattern and the simple transitive (NVN) pattern. There are also two diffuse clusters corresponding to V[NP]'s followed by *that* and S-complement verbs followed by an unambiguous determiner. The locations of these diffuse clusters reflect the attractor influences which give rise to the predictions we are interested in. The V[NP]-*that* cluster corresponds to processor states that are, from the standpoint of the symbolic grammar, equivalent to the states corresponding to the points in the V[NP]-determiner cluster. Nevertheless, this V[NP]-*that* cluster is drawn over toward the V[Sbar]-*that* cluster because of the strong influence of the V+*that*-as-complementizer attractor. This displacement of the V[NP]-*that* cluster makes the density around its points low compared to the density in the V[NP]-determiner cluster so gravitation times for points in the V[NP]-*that* region are large and we expect high latencies in reading. This is how the model predicts the results of Juliano and Tanenhaus's Experiment 2.

In fact, structural influence effects along the lines of those observed in Experiment 2 were what led researchers to propose the two-stage model. But in this case, it's not clear why processing the word *that* after a transitive verb should be difficult under the standard two-stage model which takes the NVN structure as the first-pass assumption—after all, *that* is a legitimate determiner, and thus does not conflict grammatically with the first-pass hypothesis. Moreover, the full range of data observed in these types of sentences is inconsistent with two-stage model's lack of sensitivity to lexical differences: Juliano and Tanenhaus (1994) showed that replacing the transitive verb with a strong sentence-complement verb like *insisted* makes the difficulty at the word *that* disappear.

These results are also difficult to explain under a model based on conditional probabilities generated by a reasonable linguistic grammar (see Charniak 1993, Jurafsky 1996)[1]. Reading times in such models are most naturally modeled as the uncertainty of probability distributions conditioned on grammatical context. Such models, if they calculate the probabilities accurately, can be thought of as ultra lexically-sensitive. In the present case, a pure transitive verb provides clear information that a complementizer interpretation is out of the question for *that* so the conditional probability models have no reason to predict greater difficulty with *that* than with *those* in such a context.

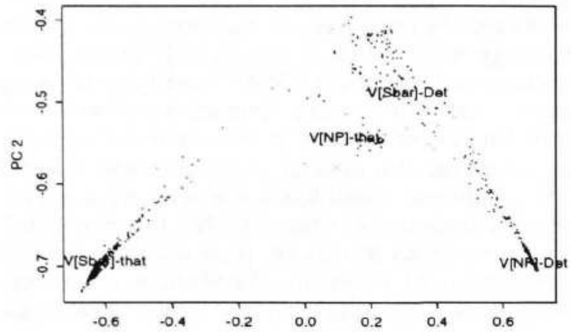
Turning now to Juliano and Tanenhaus's Experiment 3, similar attractor effects explain the difficulty of processing a unambiguous determiner like *the* after a sentence complement verb. In particular, the attraction of the V[Sbar]-determiner cluster (which is associated with a Sentence-Complement interpretation) over in the direction of the V[NP]-determiner cluster makes the density of points in the V[Sbar]-Determiner cluster low so gravitation times are high and high latencies are expected (Figure 5). Again, the two-stage model has trouble predicting this effect both because the determiner is in keeping with the preferred transitive structure and because of the sensitivity to lexical differences evidenced by the lower reading time at *the* after a transitive verb. The effect may be attributable to a statistical difference in the likelihood of *the*

[1] We note that Jurafsky's models and some of the models described by Charniak do not compute exact conditional probabilites but rather approximations. They are forced to do this for lack of sufficient data even in a very large corpus. The approximations they choose may enhance the ability of their models to handle structural influence effects like those described here by weakening the sensitivity to lexical differences. We suspect that under this weakening, their models become similar to the dynamically-interpreted neural network we describe here. We feel the dynamical analysis approach is preferable in light of the current data because it provides an explicit characterization of the structural influences in terms that can be related to known linguistic constructs.

after the two types of verbs, in which case conditional probability models may make the right qualitative prediction, but they have no capability of attributing the influence to adverse structural effects.

As noted above, to make specific reading time predictions, we used density to estimate gravitation times. For each point in the 5-dimensional hidden unit space, we chose a small radius around it[2] and counted the number of visited points within that radius. We measured the predicted reading time using Equation 4. The resulting reading time predictions are shown graphically in Figures 7 and 6. Encouragingly, the model predicts anomalies in essentially the same places as they occur for human subjects.[3]

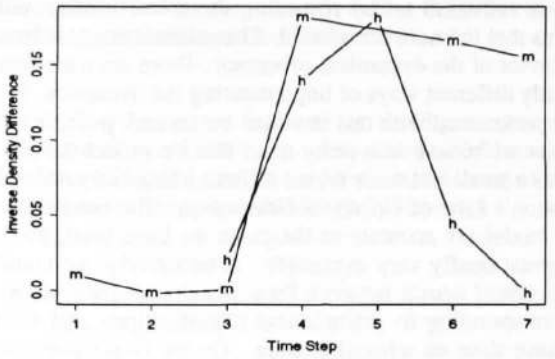
Figure 5: Hidden unit representations with cluster means labelled. (PC = Principal Component)



Conclusion

The results of the simulations are encouraging because they suggest theoretical framework for resolving the long-standing dispute in psycholinguistics over whether there is a blind syntactic processing stage. In keeping with the scheme outlined by McClelland, St. John, and Taraban 1989, the answer indicated here is that people commit (to a degree) to syntactic hypotheses as the evidence warrants but the particular commitments made are a function of accumulating lexical indications and do not reflect context-independent defaults. The capability of a wrong hypothesis to pull the processor away from the representation space location corresponding to the correct hypothesis is due to attractor competition and gives rise to the behavior pattern that has led people to posit an initial, lexically-blind processing stage. If the incorrect hypothesis/attractor is so strong that all grammatically similar cases get pulled to within a very small radius of the attractor, then no sensitivity to lexical differences is observed. But, in many cases the pulling power is of intermediate strength. In such cases lexical differences can modulate people's propensities to choose the wrong parse. Moreover, adverse attractor influences can result in high processing latencies even when the correct parse is ultimately chosen.

Figure 6: Comparison of model's predicted reading-time differences ("m") and scaled human subject data ("h") for Juliano and Tanenhaus (1993), Experiment 2. The human subject data are scaled linearly so they fall in the same range as the model's predicted reading times. (Effect is due to the V-*that*[Complementizer] attractor).



	1	2	3	4	5	6	7
	Det	Noun	Verb	Det	(Adj)	Noun	Period
(a)	the	woman	visited	that	(cheap)	hotel	p...
(b)	the	woman	visited	those	(cheap)	hotel	p...

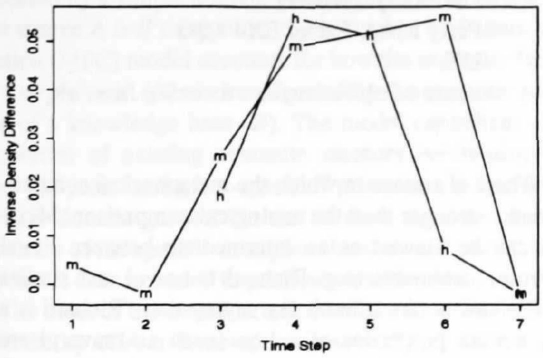
The proposal we make here is very much in keeping with the model of sentence processing proposed by McClelland, St. John, and Taraban (1989). Those authors trained a connectionist network to generate appropriate assignments of constituents to roles as a sentence was being processed. Their model showed some of the same ambivalence in the presence of ambiguous information that our model shows. Although they did not use a dynamical system to probe the representation produced by their network like we do, they are also working with a learning model which is governed by attractor dynamics, so it is likely that the predictions made in both cases have a similar source. The usefulness of introducing dynamical systems analysis to this domain is that it provides a way of identifying the structural entities in connectionist models that are responsible for those abstract constraints on language that are referred to as "syntactic". Without such tools for talking about the nature of synactic organization in connectionist models, it is hard to elucidate the relationship between those models and the standard models which make central reference to syntactic structures.

The attractor-based interpretation makes it clear why neither two-stage models nor conditional probability models based on linguistic grammars can handle all the data. In essence, the two-stage approach attributes too much respon-

[2] 0.06 hidden unit units—roughly one third the minimum distance between cluster means in the two-dimensional subspace where determiners and complementizers are distinguished. This seemed like an appropriate region in which to estimate this minimum since some of the subtlest distinctions are made in this region.

[3] The persistence of high predicted latencies at the word *hotel* and the period following *the woman visited that hotel*... in Figure 6 is due to the persisting influence of the V-*that*[Complementizer] attractor as evidenced by the fact that the model shows a nontrivial tendency to predict a verb instead of a period at the 7th timestep. Further training of the model brings this case into line with the human subject data, but tends to diminish the salience of the attractor effects in the plot corresponding to Figure 5 so we decided to use this case as an illustration despite its imperfections.

Figure 7: Comparison of model's predicted reading-time differences ("m") and scaled human subject data ("h") for Juliano and Tanenhaus (1993), Experiment 3. The human subject data are scaled linearly so they fall in the same range as the model's predicted reading times. (Effect is due to the V-NP[Direct Object] attractor).



	1 Det	2 Noun	3 Verb	4 Det	5 (Adj)	6 Noun	
(a)	the	woman	insisted	the	(cheap)	hotel	closed...
(b)	the	woman	visited	the	(cheap)	hotel	p...

siblity to structural constraints, while the conditional probability approach is too lexically sensitive.[4] Neither of these approaches allows incompatible structural constraints to interact at a given point in processing. By contrast, the attractor model succeeds by permitting such interaction to occur in a limited way. It holds, in effect, that processing is largely rule-governed, for it is dominated most of the time by single attractors which correspond to absolute interpretations. But it is marginally subject to un-rule-like influences. These occur when competing hypotheses have equal-enough sway that subtler, lexical influences can win the day.

Acknowledgements

We thank Robert Jacobs, Dan Jurafsky, Christopher Duncan and an anonymous reviewer for helpful comments. This research was supported by NICHD grant HD-27206 and NIDCD grant 5T32DC0035. Some of the data from Experiments 1-3 was presented in the 1993 Proceedings of the Cognitive Science Society Meeting.

References

Abraham, R. H. and Shaw, C. D. (1984). *Dynamics—the Geometry of Behavior, Books 0 - 4*. Aerial Press, Inc., P.O. Box 1360, Santa Cruz, CA. Volume 1 of the Visual Mathematics Library.

Charniak, E. (1993). *Statistical Language Learning*. MIT Press, Cambridge, Massachusetts.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225.

Frazier, L. (1987). Theories of sentence processing. In Garfield, J. L., editor, *Modularity in knowledge representation and natural-language understanding*, pages 37–62. MIT Press, Cambridge, MA.

Jolliffe, I. T. (1986). *Principal component analysis*. Springer-Verlag, New York.

Juliano, C. and Tanenhaus, M. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Proceedings of the 15th Annual Cognitive Science Conference*, pages 593–598. Lawrence Erlbaum.

Juliano, C. and Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research*, 23(6):459–471.

Jurafsky, D. (1996). Conditional probabilities for linguistic access and disambiguation. Paper presented at the 1996 CUNY Sentence Processing Conference, New York, New York.

MacDonald, M. A., Pearlmutter, N. J., and Seidenberg, M. S. (1994a). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101:676–703.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994b). Syntactic ambiguity resolution as lexical ambiguity resolution. In C. Clifton, J., Frazier, L., and Rayner, K., editors, *Perspectives on Sentence Processing*, pages 123–154. Lawrence Erlbaum Associates.

McClelland, J., John, M. S., and Taraban, R. (1989). Comprehension: A parallel distribute processing approach. In *Language and Cognitive Processes*, volume 4 (Special Issue), pages 287–335. Lawrence Erlbaum Associates.

Pearlmutter, B. A. (1995). Gradient calculations for dynamic recurrent networks: A survey. *IEEE Transactions on Neural Networks*, 6(5):1212–1228.

Strogatz, S. (1994). *Nonlinear Dynamics and Chaos*. Addison-Wesley, Reading, MA.

Trueswell, J. C. and Tanenhaus, M. K. (1994). Toward a constraint-based lexicalist approach to syntactic ambiguity resolution. In C. Clifton, L. Frazier, . K. R., editor, *Perspectives on sentence processing*. Erlbaum, Hillsdale, NJ.

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden paths. *Journal of Experimental Psychology*, 19(3):528–553.

Zipf, G. K. (1943). *Human Behavior and the Principle of Least Effort*. Hafner [1965].

[4] Again, we note that this lexical sensitivity can be reduced by various smoothing techniques and urge a clarification of the way the structural constraints interact via smoothing.