Robust Semantic Systematicity and Distributed Representations in a Connectionist Model of Sentence Comprehension

Stefan L. Frank (s.frank@nici.ru.nl)
Willem F. G. Haselager (w.haselager@nici.ru.nl)
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen; P.O. Box 9104
6500 HE Nijmegen, The Netherlands

Abstract

There is considerable debate about the amount and kind of systematicity displayed by neural networks. One important requirement is that the systematicity is robust and not overly dependent on specific parameter settings and training details. We present a recurrent neural network model of sentence comprehension that realizes robust semantic systematicity on the basis of distributed representations.

Introduction

Ever since Fodor and Pylyshyn (1988) claimed that connectionist models could not display systematicity and productivity, many researchers (e.g., Pollack, 1990; Chalmers, 1990; Christiansen & Chater, 1994; Niklasson & Van Gelder, 1994; Bodén & Niklasson, 2000) have attempted to provide counterexamples. In all cases, however, doubts have been raised (e.g., Haselager & Van Rappard, 1998; Aizawa, 2003; Hadley, 2004) about what exactly the proposed networks achieved. Basically, three types of criteria can be discerned in the evaluation of connectionist models that allegedly display systematicity: robustness, applicability on a larger scale, and usability in inference. Simply put, these criteria come down to the requirement that systematicity should be displayed by a network under a wide range of parameter settings and a great variety of learning circumstances, while it is dealing with a large number of (combinations of) representations, in a way that is usable for reasoning.

According to Hadley (1994a), the degree of systematicity a system displays equals the degree of input novelty it can tolerate. For a neural network, this depends on its ability to generalize from the training examples to novel (test) inputs. Hadley (1994a) reviewed several connectionist models of sentence processing and argued that none of these got close to the degree of systematicity displayed by people, which he referred to as *semantic systematicity* (Hadley, 1994b, 2004): the ability to assign a correct semantic representation to any novel sentence.

In this paper we will discuss a new model, show that it displays semantic systematicity, and evaluate it from the perspective of the criterion of robustness. It is possible to trace back this criterion to the claim made by Fodor that it is a *law* that cognitive capacities are systematic (Fodor & Pylyshyn, 1988, p. 48; Fodor & McLaughlin, 1990, pp. 202–203). According to Fodor, it is not enough to merely show that systematicity is possible, one has to indicate that it is a necessary consequence of the network's architecture. By analogy, Butler (1993, p. 232) noted that a theory of planetary motion should not merely allow for the possibility of elliptical orbits of planets, but should have the elliptical orbits as its necessary consequence. Even though we agree with several authors (e.g., Dennett, 1991, p. 27; Wilks, 1990, p. 331) that systematicity is not as 'lawful' as Fodor may suggest, we do think that Fodor's requirement is important in the sense that it provides a safeguard against too readily taking 'accidental' signs of systematicity for the real thing. After all, one would like an explanation of systematicity, not just a mere demonstration (see also Niklasson & Van Gelder, 1994, p. 297).

Even though systematicity does not have to follow necessarily out of a network's architecture, the network has to display it under a wide, and preferably psychologically realistic, set of conditions, in order to count as a genuine case of systematicity. This means that the results should not be overly dependent on the fine-tuning of parameters, nor on the specific details of carefully arranged training regimes.

One of the few robust demonstrations of semantic systematicity is formed by the sentence comprehension model of Hadley and Cardei (1999). It is different from most connectionist systems in that its processing units differ qualitatively from one another, some being 'concept nodes' while others are 'proposition nodes' or 'binding nodes'. As pointed out by Hadley, Rotaru-Varga, Arnold, and Cardei (2001, p. 74), such a model 'employ[s] *classical*, combinatorially pre-disposed, representational layers'. Much of its systematic abilities is likely to be the result of this 'combinatorially-endowed wiring'.

In contrast, the model we present here uses only genuinely distributed representations to demonstrate systematic behavior in sentence comprehension. It implements the theory that understanding a sentence requires the construction of a mental representation of the situation to which the sentence refers (cf. Zwaan & Radvansky, 1998). These situations occur in a simple 'microworld' and are described by sentences in a 'microlanguage'. In a previous version of this model (Frank, 2005), a Simple Recurrent Network (SRN; Elman, 1990) was trained to transform some of the microlanguage sentences into distributed representations of the corresponding situations. Although Frank's (2005) primary intention was not to demonstrate the network's ability to behave systematically, he did show that it could correctly process novel sentences, even if these described microworld situations on which it was not trained.

Here, the same microworld and representations (described below) are used. The microlanguage, however, is extended to allow for a stronger demonstration of semantic systematicity. Also, we apply a new type of recurrent network that is not only easier to train than a SRN but has also been shown to outperform it on a task requiring syntactic systematicity (Frank, in press) and may be more appropriate for semantic systematicity as well.

Table 1: Ten basic situations in the microworld.

situation	meaning
BOUT	Bob is outside
JOUT	Jilly is outside
SOCCER	Bob and Jilly play soccer
HIDE	Bob and Jilly play hide-and-seek
BCOMP	Bob plays a computer game
JCOMP	Jilly plays a computer game
BDOG	Bob plays with the dog
JDOG	Jilly plays with the dog
BWIN	Bob wins
JWIN	Jilly wins

Microworld and Microlanguage

Since we are dealing with semantic (as opposed to syntactic) systematicity, arising from situational representations, it is necessary to define a world in which situations occur. Obviously, the real world (or even a substantial subset thereof) is far too complex to implement in a computational model. For this reason, we make use of the microworld developed by Frank, Koppen, Noordman, and Vonk (2003) for their Distributed Situation Space (DSS) model of knowledge-based inference in story comprehension. The microworld has only two characters, called Bob and Jilly, who can perform different actions and be in several places and states. Anything that happens to them can be expressed as a boolean combination of a small number of so-called 'basic situations'. Only ten of these, listed in Table 1, are relevant here.

Of course, not just anything is possible in the microworld. For instance, soccer and hide-and-seek are always played by Bob and Jilly together. This is why there are no basic situations in which only Bob or only Jilly plays one of these games. Also, someone can win only when SOCCER or HIDE is the case, or when both Bob and Jilly play a computer game. Furthermore, soccer can only be played when Bob and Jilly are outside, while computer games are only played inside. Naturally, Bob and Jilly cannot win simultaneously.

Any situation in the microworld can be described by a sentence from a microlanguage. Sentence comprehension is taken to be the reconstruction of the situation from the sentence that describes it. In the Frank (2005) model, there were only 15 different words, which could be combined to form 328 different sentences. Moreover, word order was not relevant to these sentences. Here, a more complex microlanguage is used: It has 20 words that can form 3558 different sentences, some of which contain a transitive verb that makes word order relevant to the sentence's meaning.

The language's lexicon contains two proper names (*bob, jilly*), one pronoun (*someone*), five nouns (*dog, soccer, hide-and-seek, computer-game, game*), five verbs (*wins, loses, beats, plays, play*), two adverbs (*inside, outside*), three prepositions (*with, to, at*), and two connectives (*and, or*). These words form sentences according to the grammar in Table 2. A few examples of sentences, and the way the corresponding microworld situations are constructed from basic situations, are shown in Table 3.

Table 2: Rewrite rules of the microlanguage. The probabilities of possible productions are indicated if they are unequal. S = sentence; NP = noun phrase; VP = verb phrase; APP = adverbial/prepositional phrase. Variable n denotes verb number (singular or plural).

Head		Production
S	$\rightarrow$	$S_{play} \mid S_{win}$
S_{play}	$\rightarrow$	$NP_{sing} VP_{play,sing}$ (.7) $\mid$ $NP_{plu} VP_{play,plu}$ (.3)
S_{win}	$\rightarrow$	$NP_{sing} VP_{win,intrans} \mid NP_{sing} VP_{win,trans}$
NP_{sing}	$\rightarrow$	*someone* (.2) $\mid$ *bob* (.3) $\mid$ *jilly* (.3) $\mid$ *bob or jilly* (.1) $\mid$ *jilly or bob* (.1)
NP_{plu}	$\rightarrow$	*bob and jilly* $\mid$ *jilly and bob*
$VP_{play,n}$	$\rightarrow$	$V_{play,n}$ (.36) $\mid$ $V_{play,n}$ Place (.24) $\mid$ $V_{play,n}$ Activities (.24) $\mid$ $V_{play,n}$ Activities Place (.16)
$VP_{win,intrans}$	$\rightarrow$	$V_{win,intrans}$ (.3) $\mid$ $V_{win,intrans}$ APP (.7)
$VP_{win,trans}$	$\rightarrow$	$V_{win,trans} NP_{sing}$ (.3) $\mid$ $V_{win,trans} NP_{sing}$ APP(.7)
APP	$\rightarrow$	Place (.3) $\mid$ *at* Games (.3) $\mid$ Place *at* Games (.2) $\mid$ *at* Games Place (.2)
$V_{play,sing}$	$\rightarrow$	*plays*
$V_{play,plu}$	$\rightarrow$	*play*
$V_{win,intrans}$	$\rightarrow$	*wins* $\mid$ *loses*
$V_{win,trans}$	$\rightarrow$	*beats* $\mid$ *loses to*
Place	$\rightarrow$	*inside* $\mid$ *outside*
Activities	$\rightarrow$	*game* (.15) $\mid$ Activity (.45) $\mid$ Activity$_i$ *or* Activity$_j$ ($i \neq j$) (.4)
Activity	$\rightarrow$	Game (.75) $\mid$ *with dog* (.25)
Games	$\rightarrow$	*game* (.15) $\mid$ Game (.45) $\mid$ Game$_i$ *or* Game$_j$ ($i \neq j$) (.4)
Game	$\rightarrow$	*soccer* $\mid$ *hide-and-seek* $\mid$ *computer-game*

Note that the microlanguage (as any natural language) can describe situations that are impossible to occur, such as Jilly losing to herself or playing soccer inside. To prevent the occurrence of a very large number of such sentences, plural noun phrases and the phrase *with dog* are not allowed in sentences describing winning or losing. As a result, there are no sentences stating that both Bob and Jilly win or lose, nor any sentence stating that someone wins or loses at playing with the dog. Both these situations are impossible in the microworld.

The Model

Representing microworld situations

According to Hadley (2004, p. 150), for a sentence-comprehension model to display semantic systematicity, it is necessary that 'the resultant meaning representation for the entire sentence possesses properties which would enable us to justly claim that the entire representation could, in principle,

Table 3: Examples of microlanguage sentences and the construction of the corresponding situations from the basic situations in Table 1.

Sentence	Situation
jilly plays	SOCCER ∨ HIDE ∨ JCOMP ∨ JDOG
bob plays game	SOCCER ∨ HIDE ∨ BCOMP
bob and jilly play with dog	BDOG ∧ JDOG
someone wins	BWIN ∨ JWIN
bob or jilly wins at computer-game	(JWIN ∧ JCOMP) ∨ (BWIN ∧ BCOMP)
bob loses	JWIN
jilly loses to jilly inside	BWIN ∧ JWIN ∧¬JOUT

constrain the set of situations which could render the sentence true.' The representations used by the DSS model (which are also used here) have exactly these properties. Any situation that can occur in the microworld is represented distributively by a vector in a high-dimensional 'situation space'. As formalized below, relations among these 'situation vectors' correspond to probabilistic relations among the represented microworld situations.

Situation vectors are developed by training a Self-Organizing Map (SOM; Kohonen, 1995) on descriptions of situations occurring in the microworld (see Frank et al., 2003, for details). These descriptions take the form of binary vectors containing a 1 for each basic situation that occurs at a certain moment in time, and a 0 for each basic situation that does not. As a result of training, a membership value $\mu_i(p) \in [0,1]$ is associated to each SOM-cell i and basic situation p. This value indicates the extent to which cell i forms part of the representation of p. The SOM has 150 cells, so the representation of p can also be viewed as a 150-element situation vector of membership values $\mu(p) = (\mu_1(p),\ldots,\mu_{150}(p))$.

Representations of negations, conjunctions, and disjunctions are constructed as is common in fuzzy logic:

$$
\begin{aligned}
\mu_i(\neg p) &= 1 - \mu_i(p) \\
\mu_i(p \wedge q) &= \mu_i(p)\mu_i(q) \\
\mu_i(p \vee q) &= \mu_i(p) + \mu_i(q) - \mu_i(p)\mu_i(q),
\end{aligned}
\tag{1}
$$

where p and q can themselves be combinations of (basic) situations.

Let vector $(x_1,\ldots,x_{150})$ represent some microworld situation X.[1] The a priori probability that this situation occurs can be estimated by

$$
\tau(X) = \frac{1}{150}\sum_i x_i.
\tag{2}
$$

The content of situation X can be extracted by comparing its representation $(x_1,\ldots,x_{150})$ to several known situation vectors $\mu(p)$. From Equations 1 and 2 it follows that the conditional probability that some p is the case in situation X, can

be estimated by:

$$
\tau(p|X) = \frac{\tau(p \wedge X)}{\tau(X)} = \frac{\sum_i \mu_i(p)x_i}{\sum_i x_i}.
\tag{3}
$$

These τ-values are called *belief values* because they indicate the extent to which the DSS model may 'believe' particular (basic) situations to be the case given a situation vector. As shown empirically by Frank et al. (2003), belief values are accurate estimates of (un)conditional probabilities in the microworld. This proves that relations among microworld situations are indeed (implicitly) encoded in the organization of situation space.

The Network

A new approach to dynamical computation by neural networks was recently developed independently by Maass, Natschläger, and Markram (2002) and by Jaeger (2003). Their so-called Liquid State Machines (LSM; Maass et al.) and Echo State Networks (ESN; Jaeger)[2] are both based on the insight that training all connections of a recurrent network is not needed. Instead, the weights in the recurrent part of the network may remain fixed, greatly increasing training efficiency. These recurrent connections contribute to the network's computations by forming a 'dynamical reservoir' (DR) that stores the input sequence in an unstructured manner (in much the same way that the pattern of waves in a bucket of water contains information about what has recently fallen in). A separate 'readout' network is trained to convert the activation patterns in the DR into target outputs.

The network used here (drawn schematically in Figure 1) is in fact an extension of the ESN in that the readout network is a two-layer feedforward network, that is, it has a hidden layer. Frank (in press) reports that this additional layer improves generalization in a sentence-processing task. The complete network consists of four layers:

- The input layer, with one unit for each of the 20 words in the microlanguage.

- The dynamical reservoir, serving as a short-term memory for retaining the input sequence. The number of DR-units was varied from 80 to 250.

- The hidden layer, varied in size from 10 to 40 units.

- The output layer, with one unit for each of 150 dimensions of situation space.

Words enter the network one at a time by activating only the corresponding unit of the input layer. This activation is sent to the DR, which, like the recurrent layer of a SRN, also receives its own activation that resulted from processing the previous word. The main difference with a SRN is that the

---

[1]This can be *any* vector in situation space, that is, it does not need to have been constructed using Equation 1.

[2]Although LSMs and ESNs are conceptually similar, there are some differences in their technical descriptions and in the way they are commonly applied. The network we use here is most like an ESN, so we shall refer to it as such.
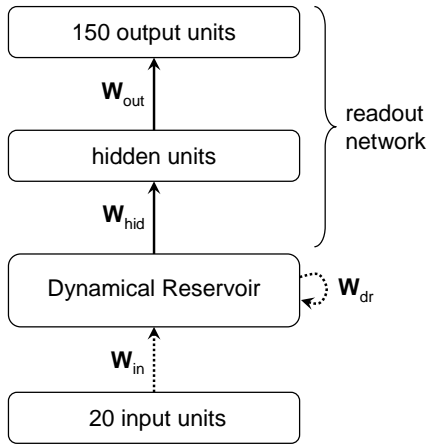
Figure 1: Extended Echo State Network. Solid arrows indicate connections with trainable weight, dashed arrows are connections with fixed weight. Connection weight matrices are denoted $\mathbf{W}$.

ESN's input and DR-weight matrices $\mathbf{W}_{in}$ and $\mathbf{W}_{dr}$ remain fixed at their initial values.[3] If the two weight matrices of the readout network, $\mathbf{W}_{hid}$ and $\mathbf{W}_{out}$, are trained correctly, the network's output activation vector represents the microworld situation referred to by the input sentence.

Although $\mathbf{W}_{dr}$, the matrix of DR connection weights, has random values, not just any matrix will do. First, it is important that only a small fraction of possible DR-connections is present. In the network used here, 85% of DR-connection weights are set to zero. Second, the optimal overall scaling of weights depends on the network's task. This scaling is expressed by the spectral radius of $\mathbf{W}_{dr}$,[4] which was varied from .4 to .95. The larger this value, the longer information remains in the DR.

Training and Testing

The network was trained on a set of 10 000 sentence tokens, randomly generated by the production rules of Table 2. The following sentences were excluded from the training set:

- Those with *bob*, *bob or jilly*, or *jilly or bob* as object of *beats*.

- Those with *jilly*, *bob or jilly*, or *jilly or bob* as object of *loses to*.

- Those containing both *dog* and *inside*.

- Those containing both *hide-and-seek* and *outside*.

This means that the network does not learn to construct any situation in which Bob and/or Jilly plays with the dog inside, nor any situation in which they play hide-and-seek outside.

---

[3] Furthermore, in our network, DR-units are linear and have no bias weights, while all units of a SRN usually have sigmoidal activation functions and bias weights.

[4] The spectral radius of a matrix is its largest absolute eigenvalue.

The network *is* trained on situations in which Bob loses or Jilly wins. However, these situations are only described using the intransitive verbs *wins* and *loses*, not the transitive *beats* and *loses to*.

The microworld situation referred to by each training sentence was the target output during training on that sentence. After each word of each training sentence, connection weights $\mathbf{W}_{hid}$ and $\mathbf{W}_{out}$ were updated using the standard backpropagation algorithm.

After training, the network was tested on three sets of test sentences:

- The 'win'-set, consisting of all 88 sentences that describe a possible situation in the microworld, do not specify a place, and have either *bob* as object of *beats* or *jilly* as object of *loses to*.

- The 'game/place'-set, consisting of all 14 sentences that describe a possible situation in the microworld, and end in either *play(s) with dog inside* or *play(s) hide-and-seek outside*.

- The 'random'-set, consisting of 100 random sentences.

The sentences in the 'random'-set may or may not have been training sentences. In contrast, the 'win' and 'game/place'-sets contain only novel sentences. Moreover, the *situations* described by the 'game/place'-sentences were not presented to the network during training. To successfully process these sentences, the network must have learned how novel combinations of known words refer to novel combinations of known situations. That is, it must display semantic systematicity.

Rating performance

Comprehension is not an all or none type of phenomenon. Millis, King, and Kim (2000), for instance, showed experimentally that the extent to which readers develop a situational representation of a text varies between subjects and experimental tasks. The model, too, may only construct the correct situation vector to a degree. To ascertain the network's performance, the comprehension measure defined by Frank (2005) is used. It is based on the belief values from Equations 2 and 3.

Suppose the network processes a sentence that states p, and the resulting output vector represents some situation X. The associated comprehension score is then defined as

$$\frac{\tau(p|X) - \tau(p)}{\tau(p|p) - \tau(p)}.$$

If the probability of p in situation X is larger than its a priori probability (i.e., if $\tau(p|X) > \tau(p)$), processing the sentence has increased belief in p. This means that the sentence was understood to some extent, as indicated by a positive comprehension score. A perfect network will result in $X = p$, so $\tau(p|X) = \tau(p|p)$, corresponding to a comprehension score of 1. In contrast, if processing the sentence *decreases* belief

in p (i.e., if $\tau(p|X) < \tau(p)$), the sentence was misunderstood and the network made an error. The resulting comprehension score is negative.

Most sentences describe more than one basic situation. For instance, *jilly wins at hide-and-seek outside* describes a game (Jilly plays hide-and-seek), a place (Jilly is outside) and an outcome (Jilly wins). To investigate whether the network understood all of these basic situations, comprehension scores are computed not only for the complete situation HIDE $\wedge$ JOUT $\wedge$ JWIN, but also separately for the three basic situations. In general, each sentence may give rise to up to four separate comprehension scores: for the game, the place, the outcome, and the conjunction of the three.

Results

To investigate the robustness of the training process and the model's results, three parameters were varied: the number of DR-units ($80 - 250$), the number of hidden units ($10 - 40$), and the spectral radius of $\mathbf{W}_{dr}$ ($.4 - .95$).

Apart from the smallest networks, all combinations of parameter settings resulted in positive comprehension scores for all test sets (averaged over test sentences). In general, the larger the network, the better its performance. For the largest network, the optimal spectral radius was around .6.

Applying this optimal parameter setting, ten networks were trained, differing only in their initial random weights. The resulting comprehension scores, averaged over the ten networks and over test sentences, are listed in Table 4. Clearly, all scores are significantly positive, indicating that the network understood each part of the sentences (as well as complete sentences) of each test set. The table also shows the percentage of errors. Most of these are very low, meaning that only a few sentences were misunderstood.

Conclusions

The sentence comprehension model clearly behaves systematically: It was not trained on sentences stating that *jilly beats bob* or *bob beats jilly*, yet it could correctly construct representations of these situations. Furthermore, it was never trained to represent situations in which Bob and/or Jilly play with the dog inside, nor situations in which hide-and-seek is played outside. Nevertheless, it did comprehend sentences referring to such situations.

These results are quite robust in the sense that they did not crucially depend on a delicate setting of parameters, nor required an extensive and complex training procedure. Ten networks were trained, starting with different weight settings, and each of these managed to comprehend the test sentences. Also, we demonstrated that many novel sentences were comprehended correctly, again indicating that the network's abilities are not just accidental.

As stated in the Introduction, two more criteria are often applied to evaluate systematicity in connectionist models: scalability and usability in inference. It is only fair to admit that the size and complexity of the simulations need to be expanded considerably in order to be psychologically realistic. Although the microlanguage was 10 times the size of Frank's (2005b), other models of language processing have used more extensive languages. However, it should be kept in mind that our focus on microworld situations requires the time-consuming and effortful construction of a microworld and corresponding microlanguage, making large-scale simulations laborious to set up. As for the criterion of usability in inference, be reminded that the situational representations were developed for the DSS model of inferencing in story comprehension (Frank et al., 2003). Correct predictions of experimental data by the DSS model indicate that these representations not only allow for simulations of inferencing, but do so in a realistic manner.

Acknowledgments

References

Aizawa, K. (2003). *The systematicity arguments.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bodén, M., & Niklasson, L. (2000). Semantic systematicity and context in connectionist networks. *Connection Science*, *12*, 111-142.

Butler, K. (1993). Connectionism, classical cognitivism and the relation between cognitive and implementational levels of analysis. *Philosophical Psychology*, *6*, 321–333.

Chalmers, D. J. (1990). Syntactic transformations on distributed representations. *Connection Science*, *2*, 53–62.

Christiansen, M. H., & Chater, N. (1994). Generalization and connectionist language learning. *Mind & Language*, *9*, 273–287.

Table 4: Average comprehension scores with 95% confidence intervals, and percentage of errors for all three test sets.

set	part	comprehension	% errors
win	game	$.80 \pm .02$	0.0
	outcome	$.30 \pm .01$	0.0
	complete	$.28 \pm .01$	0.0
game/place	game	$.29 \pm .03$	7.1
	place	$.76 \pm .03$	0.0
	complete	$.27 \pm .02$	0.0
random	game	$.77 \pm .02$	2.9
	outcome	$.49 \pm .03$	0.4
	place	$.70 \pm .04$	7.4
	complete	$.60 \pm .02$	2.2

Dennett, D. C. (1991). Mother nature versus the walking encyclopedia: a western drama. In W. Ramsey, S. Stich, & D. Rumelhart (Eds.), *Philosophy and connectionist theory*. Hillsdale, NJ: Erlbaum.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fodor, J. A., & McLaughlin, B. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution does not work. *Cognition*, *35*, 183–204.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, *28*, 3–71.

Frank, S. L. (2005). Sentence comprehension as the construction of a situational representation: a connectionist model. In A. Russell, T. Honkela, K. Lagus, & M. Pöllä (Eds.), *Proceedings of AMKLC'05*. Espoo, Finland: Helsinki University of Technology.

Frank, S. L. (in press). Learn more by training less: systematicity in sentence processing by recurrent networks. *Connection Science*.

Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, *27*, 875–910.

Hadley, R. F. (1994a). Systematicity in connectionist language learning. *Mind & Language*, *9*(3), 247–272.

Hadley, R. F. (1994b). Systematicity revisited: reply to Christiansen and Chater and Niklasson and van Gelder. *Mind & Language*, *9*, 431–444.

Hadley, R. F. (2004). On the proper treatment of semantic systematicity. *Minds and Machines*, *14*, 145–172.

Hadley, R. F., & Cardei, V. C. (1999). Language acquisition from sparse input without error feedback. *Neural Networks*, *12*, 217–235.

Hadley, R. F., Rotaru-Varga, A., Arnold, D. V., & Cardei, V. C. (2001). Syntactic systematicity arising from semantic predictions in a Hebbian-competitive network. *Connection Science*, *13*(1), 73–94.

Haselager, W. F. G., & Van Rappard, J. F. H. (1998). Connectionism, systematicity, and the frame problem. *Minds and Machines*, *8*, 161–179.

Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (Vol. 15). Cambridge, MA: MIT Press.

Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, *14*, 2531–2560.

Millis, K. K., King, A., & Kim, H. J. (2000). Updating situation models from descriptive texts: a test of the situational operator model. *Discourse Processes*, *30*, 201–236.

Niklasson, L. F., & Van Gelder, T. (1994). On being systematically connectionist. *Mind & Language*, *9*, 288–302.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence*, *46*, 77–105.

Wilks, Y. (1990). Some comments on Smolensky. In D. Patridge & Y. Wilks (Eds.), *The foundations of AI*. Cambridge: Cambridge University Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation model in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.