Integrating Multiple Cues in Word Segmentation: A Connectionist Model using Hints

Joe Allen and Morten H. Christiansen
Program in Neural, Informational and Behavioral Sciences
University of Southern California
University Park MC-2520
Los Angeles, CA 90089-2520
joeallen@gizmo.usc.edu morten@gizmo.usc.edu

Abstract

Children appear to be sensitive to a variety of partially informative "cues" during language acquisition, but little attention has been paid to how these cues may be integrated to aid learning. Borrowing the notion of learning with "hints" from the engineering literature, we employ neural networks to explore the notion that such cues may serve as hints for each other. A first set of simulations shows that when two equally complex, but related, functions are learned simultaneously rather than individually, they can help bootstrap one another (as hints), resulting in faster and more uniform learning. In a second set of simulations we apply the same principles to the problem of word segmentation, integrating two types of information hypothesized to be relevant to this task. The integration of cues in a single network leads to a sharing of resources that permits those cues to serve as hints for each other. Our simulation results show that such sharing of computational resources allows each of the tasks to facilitate the learning (i.e., bootstrapping) of the other, even when the cues are not sufficient on their own.

Introduction

A theory language of acquisition requires an explanation for how and why children learn the complexities of their native languages so quickly, effortlessly and uniformly. Most traditional answers to this question have taken the form of claims that children are born with language specific constraints, in part because of a gap between the input to which the child is exposed and the competence later exhibited by the adult. The problem, as traditionally characterized, is that the data alone are insufficient to determine the nature of the underlying system, and that therefore additional sources of information are necessary for successful acquisition to occur. Interestingly, a very similar problem has been faced by the engineering oriented branch of the neural network community, in which the problem is construed as learning a function from a limited set of examples. From these investigations have emerged a number of alternative methods for incorporating information not present in the example set into the learning process. These additional sources of information, many based on non-intuitive properties of neural networks, have come to be referred to as "hints". In this paper, we present a novel way of looking at learning with hints within the setting of connectionist modeling of language.

Hints facilitate learning by reducing the number of candidate solutions for a given task (Abu-Mostafa, 1990) and have been shown to result in better generalization (Al-Mashouq & Reed, 1991; Suddarth & Kergosien, 1991) as well as faster learning (Abu-Mostafa, 1990; Al-Mashouq & Reed, 1991; Gällmo & Carlström, 1995; Omlin & Giles, 1992; Suddarth & Kergosien, 1991). The introduction of hints into neural networks has taken various forms, ranging from explicit rule insertion via the pre-setting of weights (Omlin & Giles, 1992), to task specific changes in the learning algorithm (Al-Mashouq & Reed, 1991), to perhaps the most interesting kind of hint: the addition of extra *"catalyst"* output units. Catalyst units are used to represent additional target values expressing a function correlated with, but simpler than, the original target function. The use of catalyst units forces the network to find an internal representation which approximates both the target and the related catalyst function. Suddarth & Kergosien (1991) list a number of simulation experiments in which this approach resulted in faster learning and better generalization. The use of catalyst units has also found its way into engineering applications—e.g., controlling link admissions ATM telecommunication networks (Gällmo & Carlström, 1995).

The idea of inserting information into a network before training has received some attention within cognitive science (albeit not understood in terms of hints). For instance, Harm, Altmann & Seidenberg (1994) demonstrated how pretraining a network on phonology can facilitate the subsequent acquisition of a mapping from orthography to phonology (thus capturing the fact that children normally have acquired the phonology of their native language—that is, they can talk—before they start learning to read). However, catalyst hints have not been explored as a means of improving connectionist models of language. In particular, there is the possibility (not investigated in the engineering hint literature) that such hints could become more than just a catalyst; that is, there may be cases where the learning of two or more functions by the same system may be superior to trying to learn each function individually. Children appear to integrate information from a variety of sources—i.e., from multiple *"cues"*—during language acquisition (Morgan, Shi & Allopenna, 1996), but little attention has been paid to potential mechanisms for such integration. We suggest that cues may serve as "hints" for each other, in that each task constrains the set of solutions available for the other task(s).

In what follows, we show that when two related functions

are learned together each is learned faster and more uniformly. We first provide a simple illustration of the advantage of the integrated learning of two simple functions, XOR and EVEN PARITY, over learning each of them separately. Next, the same idea is applied to a more language-like task: the integrated learning of word boundaries and sequential regularities given a small vocabulary of trisyllabic nonsense words. Finally, in the conclusion, we discuss possible implications for models of language acquisition.

The integrated learning of XOR and EVEN PARITY

In order to provide a simple example of the advantage of allowing two functions to interact during learning, we carried out a series of simulations involving the two simple non-linear functions: XOR and EVEN PARITY. The XOR function has been used before to demonstrate how learning with an extra catalyst unit can decrease convergence time significantly (Gällmo & Carlström, 1995; Suddarth & Kergosien, 1991), but these studies used simpler linear functions (such as, AND) to provide hints about the more complex function. In contrast, we use two functions of equal computational complexity.

input	XOR(1)	EP(1)	XOR-EP	XOR(2)	EP(2)
0 0	0	1	1 0	0 0	1 0
1 1	0	1	1 0	0 0	1 0
1 0	1	0	0 1	0 1	0 0
0 1	1	0	0 1	0 1	0 0

Table 1: The input and required output for the five training conditions.

Given two inputs, i_1 and i_2, XOR is true (i.e., 1) when $(i_1 + i_2) \bmod 2 = 1$. EVEN PARITY is the logical negation of XOR and is true when $(i_1 + i_2) \bmod 2 = 0$ (in fact, XOR is also known as ODD PARITY). The output of the XOR and EVEN PARITY functions given the four possible binary input combinations is displayed in Table 1 as XOR(1) and EP(1), respectively. These two functions can be learned by a 2-2-1 multi-layer feedforward network. Learning XOR and EVEN PARITY simultaneously requires two output units (i.e., a 2-2-2 net), and the required output is shown as XOR-EP in Table 1. For comparison, two additional 2-2-2 nets were also trained on the individual functions from which the output is labeled XOR(2) and EP(2).

A total of 100 networks (with different initial weight randomizations) were trained for each of the five input/output combinations[1] . Figure 1 illustrates the Root Mean Square (RMS) error history as a function of the number of iterations for nets trained on the XOR-EP, XOR(1), and EP(1) training conditions. Given the assumption that a net has converged

[1]Identical learning parameters were applied in all training conditions: learning rate = .1; momentum = .95; initial weight randomization = [-.1;.1]; number of training iterations = 2000.

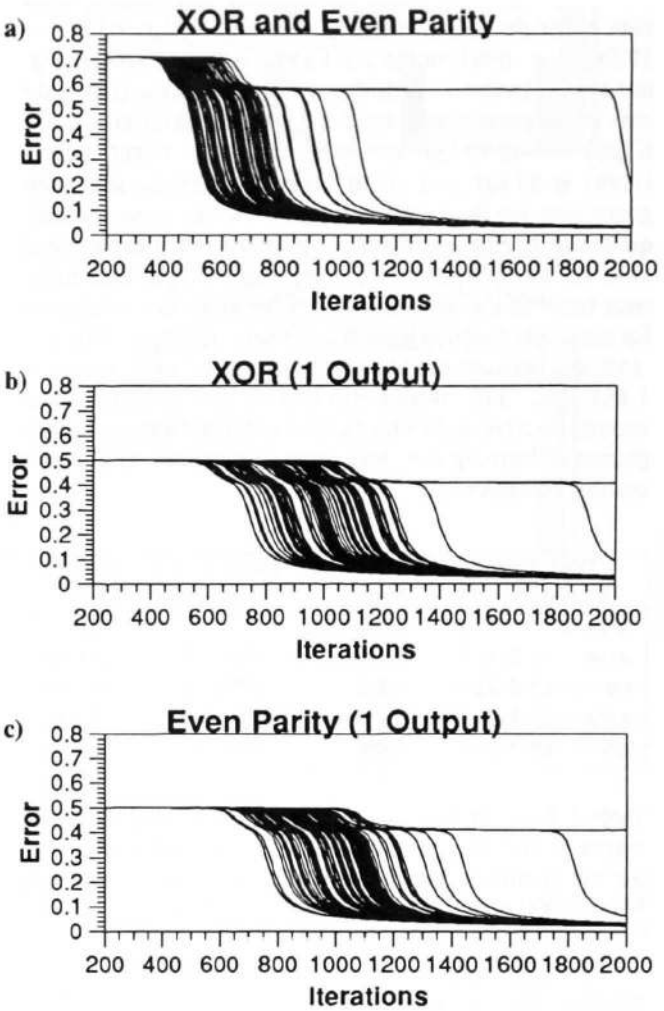

Figure 1: RMS error for 100 sessions of learning a) XOR and EVEN PARITY in a 2-2-2 net, b) XOR in a 2-2-1 net, and c) EVEN PARITY in a 2-2-1 net.

when the average RMS error for the four input/output combination is .15 or below, we can provide a quantitative analysis of the performance of the nets in the different training conditions. The results in Table 2 clearly show that the networks trained simultaneously on both XOR and EVEN PARITY (see Figure 1a) reach convergence significantly faster than either the 2-2-1 XOR nets (see Figure 1b): $t(184) = 17.589, p < .0001$; or the 2-2-1 EVEN PARITY nets (see Figure 1c): $t(184) = 20.056, p < .0001$. This is also the case in comparison with the 2-2-2 nets trained on XOR alone: $t(189) = 42.797, p < .0001$; and EVEN PARITY alone: $t(189) = 38.876, p < .0001$. Thus, the decrease in convergence time for the XOR-EP trained networks is not a consequence of having additional weights due to the extra output unit. As was to be expected, there was no significant difference between the mean number of iterations to convergence for the single function trained 2-2-1 nets: $t(184) = .610, p = .542$; and between the likewise trained 2-2-2 nets: $t(194) = 1.581, p = .115$. Notice also that the nets trained simultaneously on XOR and EVEN PARITY exhibited a more *uniform* pattern of learning (i.e., less variation) that any of the other training conditions.

Training Condition	Convergence Rate	Mean no. of Iterations	Standard Deviation
XOR + EP	93%	710.96	99.47
XOR (2-2-1)	93%	1063.87	165.95
EP (2-2-1)	93%	1077.84	145.69
XOR (2-2-2)	98%	1519.69	154.32
EP (2-2-2)	98%	1483.57	165.33

Table 2: Convergence rate, mean number of iterations to convergence, and the standard deviation of this mean for each of the five training conditions (only data from nets converging within 2000 iterations are included in the table).

The results from the above simulation experiments confirm that there are cases where the integrated learning of two functions of equal complexity is better than seeking to learn each of the functions individually. A possible explanation can be found in Suddarth & Kergosien (1991) who analyzed weight changes during the learning of XOR with and without hints. They found that hints allow networks to escape local minima positioned at the origin of weight space. What we have shown here is that a hint need not be a simpler function than the original target function. The results indicate that if two functions are equally complex, but sufficiently correlated, then it may be advantageous to have a single network learn them together. Even though XOR and EVEN PARITY are negations of each other, they are similar in that successful learning of either function requires partitioning the state space in the same way (with the input "1 0" and "0 1" being treated different from "1 1" and "0 0"). The two functions may thus help "bootstrap" each other by forcing their shared resources (in this case the hidden units) toward a common organization of the input. A mechanism which allows two or more functions to bootstrap each other is of potential relevance to the study of language acquisition since children appear to be sensitive to multiple speech cues which by themselves do not appear to be sufficient to bootstrap language. Of course, learning XOR and EVEN PARITY is a far cry from the task facing children acquiring their native language. We therefore turn to a more language-like application of the idea of bootstrapping via the integrated learning of multiple functions.

Integrating Cues in Word Segmentation

In order to understand an utterance a child must first be able to segment the speech stream into words. While it is likely that adult word level speech segmentation occurs partly as a byproduct of word recognition, infants lack the lexical knowledge which is a pre-requisite to this procedure. A number of proposals regarding bottom up exploitation of sub-lexical cues have been put forward to explain the onset of this capacity (e.g., Jusczyk, 1993). These proposals would require infants to integrate distributional, phonotactic, prosodic and rhythmic information in the segmentation process. In this connection, Brent & Cartwright (in press) have shown that a statistically based algorithm utilizing distributional regularities (including utterance boundary information) is better able to segment words when provided with phonotactic rules. Whereas the process of identifying and verifying the existence and potential of various cues is receiving considerable effort, there has been little attention paid to psychologically plausible mechanisms potentially responsible for integrating these cues. An understanding of possible integrating mechanisms is important for evaluating claims about the potential value of cues. Each of the cues to basic grammatical category measured by Morgan, Shi & Allopenna (1995), for example, had low validity with respect to distinguishing between the categories they considered, but taken together the set of cues was shown to be sufficient in principle to allow a naive learner to assign words to rudimentary grammatical categories with very high accuracy.

Previous connectionist explorations of word segmentation have mainly focused on single cues. Thus, Aslin, Woodward, LaMendola & Bever (1996) demonstrated that utterance final patterns (or boundaries) could be used by a back-propagation network to identify word boundaries with a high degree of accuracy. Cairns, Shillcock, Chater & Levy (1994), on the other hand, showed that sequential phonotactic structure could serve as a cue to the boundaries of words. In contrast, our investigation concentrates on the *integration* of these two cues to word segmentation. The purpose of our simulations is to demonstrate how distributional information reflecting phonotactic regularities in the language may interact with information regarding the ends of utterances to inform the word segmentation task in language acquisition. In particular, we apply the principle of catalyst learning to ask whether learning distributional regularities will assist in the discovery of word boundaries, and whether the learning of word boundaries facilitates the discovery of word internal distributional

regularities. As an initial hypothesis, we propose that as a property of the integrating mechanism, the language acquisition system makes use of the efficiency provided by the *sharing of resources* demonstrated in the XOR-EVEN PARITY simulations above to facilitate the task of segmenting the speech stream prior to lexical acquisition.

Saffran, Newport & Aslin (in press) show that adults are capable of acquiring sequential information about syllable combinations in an artificial language such that they can reliably distinguish words that conform to the distributional regularities of such a language from those that do not. For our simulations we constructed a language whose basic constituents were four consonants (*"p"*, *"t"*, *"b"*, *"d"*) and three vowels (*"a"*, *"i"*, *"u"*). These were used to create two vocabulary sets. The first consisted of fifteen trisyllabic words (e.g., *"tibupa"*). Because we hypothesize (like Saffran *et al.*) that variability in the word internal transitional probabilities between syllables[2] serves as an information source regarding the structure of the input language, some syllables occurred in more words, and in more locations within words, than others. Built into this vocabulary were a set of additional restrictions. For example, there are no words that begin with *"b"*, and no words ending in *"u"*. We will refer to this vocabulary set as *"vtp"* (for variable transitional probability). A *"flat"* vocabulary set, consisting of 12 items, was made up of words with no "peaks" in the word internal syllabic probability distribution; that is, the probability of a given consonant following a vowel was the same for all consonants (and *vice versa* for the vowels)[3]. The flat vocabulary set did not contain any additional restrictions.

Training corpora were created by randomly concatenating 120 instances of each of the words (in a particular vocabulary set) into utterances ranging between two and six words. An additional symbol marking the utterance boundary was added to the end of each utterance, but word boundaries were not marked. The utterances were concatenated into a single large input string. Simple recurrent networks[4] (Elman, 1990) were trained on these corpora by presenting letters one at a time. The task of the networks was to predict the next item in the string. For the testing phase, versions of the input corpora without utterance boundary markers were presented once to the networks.

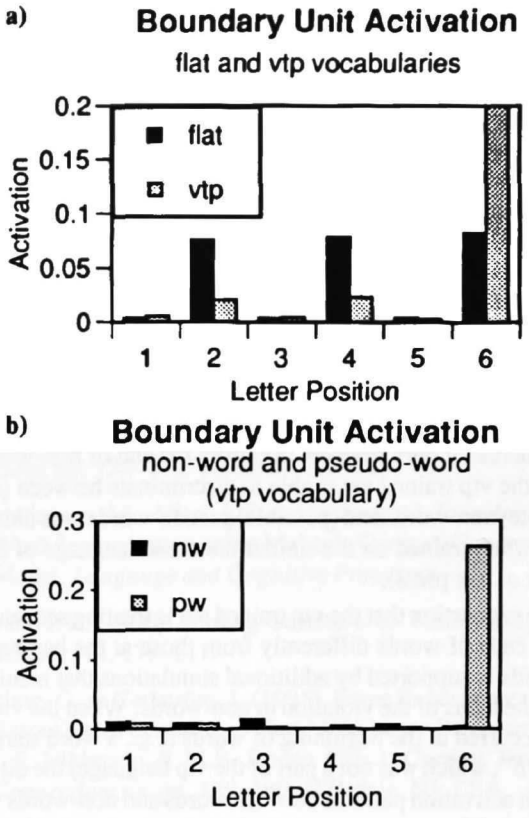

Figure 2: Average activation of the boundary unit from each letter position for a) nets trained, respectively, on the flat and the vtp vocabularies, and b) the net trained on the vtp vocabulary and tested on non-words (nw) and pseudo-words (pw).

A comparison between the networks trained respectively on the vtp and flat vocabularies is given in Figure 2(a), which shows the average activation of the boundary unit from each position[5] in a word across the two test corpora. The vtp trained network predicts a boundary with significantly higher confidence at word boundaries than at non word boundaries ($t(12598) = 87.059, p < .0001$). For the vtp trained net, average activation of the boundary unit from the ends of words was .204, while average activation of the boundary unit from positions within words was .04. The network trained on the flat vocabulary, on the other hand, shows almost no discrimination between end-of-word and non-end-of-word positions. Thus the net trained on both the vocabulary with much variation in the syllabic distributional regularities (vtp) and utterance boundary information differentiates ends of words from other parts of words, whereas a network trained only on boundary markers (and the flat vocabulary with little variation

[2]The transitional probability between syllables is defined as number of occurrences of syllable X before syllable Y in proportion to the number of occurrences of syllable X; i.e., $\frac{freq.of XY}{freq.of X}$.

[3]For the vtp vocabulary set, transitional probabilities between syllables word internally ranged from .3 to .7. The transitional probabilities between syllables across word boundaries were lower, ranging between .1 and .3. For the flat vocabulary set, the word internal transitional probabilities between syllables was .667, and did not differ from one another. The transitional probabilities across word boundaries in the full corpus ranged between .007 and .18.

[4]The network architecture consisted of 8 input/output units, (each representing a single letter, plus one representing the boundary marker), 30 hidden units and 30 context units. Identical learning parameters were applied in all training conditions: learning rate .1; momentum .95; initial weight randomization [-.25; .25]; number of training iterations = 7.

[5]Since each network is faced with a prediction task, activation from a position in Figure 2 corresponds to the network's prediction as to the next item in the string; e.g., the network's prediction for letter position 2 is plotted above letter position 1.

in syllabic probability distribution) fails to do so.

In order to assess generalization, we tested the vtp trained network on pseudo-words and non-words. For our pseudo-word trials, we presented the vtp trained network with a set of novel words that were legal in our language (e.g., *"tubipa"*). For our non-word trials, we created a set of words violating the built-in constraints of our language. In this case, we used words ending in *"u"* (e.g., *"tudadu"*). Figure 2(b) shows the average activation level of the boundary unit from each position in the non-word and pseudo-word trials. The activation of the boundary unit stays low for all positions for both types of words except from the final position of the pseudo-words, where the activation level jumps. The average activation of the boundary unit from the end of pseudo-words was .26, whereas it only reached .006 from the end of non-words. Thus, the vtp trained net is able to discriminate between (impossible) non-words and (possible) pseudo-words—as did humans when trained on the similar nonsense language of Saffran *et al.* (in press).

The suggestion that the vtp trained net is treating sequences at the ends of words differently from those at the beginning of words is supported by additional simulations that manipulated the locus of the violation in non-words. When the violation occurred at the beginning of words (e.g., a word starting with *"b"*, which was not a part of the vtp language) the difference in activation patterns between words and non-words was not as readily apparent as when the violation occurred at the ends of words. This result may correspond to the fact that human subjects in the Saffran *et al.* experiments confused legal words with non-words more often when the non-words were made up of the ends of (legal) words than when the non-words were made up of the beginnings of (legal) words.

The results presented above show that if a network learning a simple language with strong variation in the transitional probabilities between syllables has access to the additional information provided by the silences at the ends of utterances, it can use those probabilities to make better hypotheses about the locations of likely word boundaries than a network trained on a language with flat transitional probabilities between syllables. This suggests that the variability in transitional probabilities between syllables may play an important role in allowing learners to identify probable points at which to posit word boundaries. In other words, a network with access to both transitional probabilities and utterance boundary information performs better on a measure of identifying likely word boundaries than a network with access to only utterance boundary information. We can also measure the reverse, i.e., the extent to which utterance boundary information is helpful to learning distributional regularities. In order to do so, we now turn to a simulation that compares the vtp net trained with and without utterance boundary information.

The two nets were tested on a string consisting of the original 15 words in the vocabulary set (with no word or utterance boundaries marked). The test revealed only minor differences between the two networks, in all likelihood because the built-in distributional regularities are so strong in the small language as to create a ceiling effect. This interpretation is corroborated by a repetition of the experiment using the flat vocabulary: the network trained with boundary markers showed significantly better performance (measured in terms of RMS error) than that trained without boundary markers ($t(142) = 2.012, p < 0.05$). The presence of boundary markers in the input significantly altered the outcome of learning, such that the net trained with boundary markers was better able to learn the sequential regularities which were present in the flat corpus[6]. That is, the integrated learning of two functions again results in better performance. If a network has access to sequential information and utterance markers, it learns the sequential regularities better than a network with access only to sequential information. This result is consistent with the hypothesis that the silences at the ends of utterances may play an important role in the discovery of language specific phonotactic regularities.

Discussion

In the series of simulations reported here we adapted the catalyst hint mechanism previously employed in the engineering literature to the learning of two sufficiently related functions. We demonstrated that the integrated learning of two such functions may result in faster and better learning by combining the well known XOR and EVEN PARITY functions into a single 2-2-2 network. The same idea was then applied to two of the forms of information hypothesized to be relevant to the word segmentation problem by combining strongly constrained distributional information with information about the locations of utterance boundaries in a corpus of utterances generated from an artificial vocabulary of trisyllabic nonsense words. Results suggest that the simultaneous presence of both types of information in the same system may allow them to interact in such a way as to facilitate the acquisition of both phonotactic knowledge and the ability to segment speech into words.

There are several apparent differences between the XOR and EVEN PARITY simulations in section 2 and the simulations presented in section 3. First, the former simulations are of independent functions, both of which can be learned on their own without the presence of the other. The prediction of boundary markers reported in section 3, on the other hand, is not independent of the letter sequences in which they were embedded. That is, although the XOR and EVEN PARITY tasks may be learned separately, learning which of the letter sequences predicts a boundary cannot be learned independently from learning the letter sequences themselves. However, although as observers we can see XOR and EVEN PARITY as independent problems, the network, of course, does not do so. It is treating both (sub)tasks as a part of the larger task to be

[6] Although the flat vocabulary did not differ with respect to the transitional probabilities between *syllables*, the transitional probabilities between letters (and sequences longer than the syllable) did differ. We take these to be the source of regularity used by the networks in learning the structure of the flat vocabulary set.

solved. In the XOR-EP simulations, the requirements of each task constrain the solution for the other. A similar claim holds for the simulations presented in section 3. As the simulations themselves verify, these two information sources can be seen as distinct, and can be manipulated independently. But the network is treating both parts of the problem together, and shows an advantage for each task under these conditions.

Although the two sets of simulations differ in important ways, we suggest that the same mechanism is responsible for the results in both section 2 and 3. Just as XOR and EVEN PARITY can be viewed as independent problems, we can see the prediction of word boundaries as a separate task from that of predicting the next letter in the sequence. Because the tasks are learned together, the presence of a secondary task alters the solution applied to the primary task. Specifically, successfully predicting boundaries requires the network to rely on longer sequences than a network required only to predict the next letter. For example, even though consonants can be predicted largely on the basis of the preceeding letter (*"a"* implies *"b"*, *"d"*, *"t"* and *"p"* roughly equally), the end of an utterance is not predictable unless larger sequences are taken into account (e.g., *"a"* predicts an utterance boundary only when preceeded by *"ub"*, etc). The architecture of the network allows it to discover the particular distributional window by which it can perform the entire task optimally. The presence of the word boundary prediction task encourages the net to find an overall solution based on longer letter sequences, just as the presence of the XOR problem encourages the XOR-EP net to find a solution to the EVEN PARITY problem compatible with that which will solve XOR.

Although we have concentrated here on only a few sources of information relevant to the initial word segmentation problem, many additional cues to this task have been proposed (Jusczyk 1993). Our model is not, of course, meant as a complete account of the acquisition of these skills. Admittedly, prior connectionist investigations of the word segmentation problem by Aslin et al. (1996) and Cairns et al. (1994) used more realistic training samples than our artificial language. However, we have concentrated here on the advantages provided by a connectionist integration mechanism, and have successfully extended our approach to a corpus of phonetically transcribed child directed speech (Christiansen, Allen & Seidenberg, in submission). In this connection, a fundamental question for language acquisition theory is why language development is so fast, and so uniform, across children. Although most traditional answers to this question have been based on the idea that children are born with language specific constraints, the speed and uniformity provided by simultaneous learning of related functions may also provide constraints on the development of complex linguistic skills.

References

Abu-Mostafa, Y.S. (1990) Learning from Hints in Neural Networks. *Journal of Complexity*, **6**, 192–198.

Al-Mashouq, K.A. & Reed, I.S. (1991) Including Hints in Training Neural Nets. *Neural Computation*, **3**, 418–427.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996) Models of Word Segmentation in Fluent Maternal Speech to Infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax*, pp. 117–134. Mahwah, NJ: LEA.

Brent, M.R. & Cartwright, T.A. (in press) Distributional Regularity and Phonotactic Constraints are Useful for Segmentation. *Cognition*.

Cairns, P., Shillcock, R., Chater, N. & Levy, J. (1994) Lexical Segmentation: The Roles of Sequential Statistics in Supervised and Un-supervised Models. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pp. 136–141. Hillsdale, NJ: LEA.

Christiansen, M., Allen. J. & Seidenberg, M. (in submission) Word Segmentation using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*.

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

Gällmo, O. & Carlström, J. (1995) Some Experiments Using Extra Output Learning to Hint Multi Layer Perceptrons. In L.F. Niklasson & M.B. Boden (Eds.), *Current Trends in Connectionism*, pp. 179–190. Hillsdale, NJ: LEA.

Harm, M., Altmann, L., & Seidenberg, M. (1994) Using Connectionist Networks to Examine the Role of Prior Constraints in Human Learning. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pp. 392–396. Hillsdale, NJ: LEA.

Jusczyk, P. W. (1993) From general to language-specific capacities: the WRAPSA Model of how speech perception develops. *Journal of Phonetics*, **21**, 3–28.

Morgan, J., Shi, R., & Allopenna, P. (1996) Perceptual Bases of Rudimentary Grammatical Categories: Toward a Broader Conceptualization of Bootstrapping. In J. Morgan & K. Demuth (Eds.), *From Signal to Syntax*, pp. 263–281. Mahwah, NJ: LEA

Omlin, C.W. & Giles, C.L. (1992) Training Second-Order Recurrent Neural Networks Using Hints. In D. Sleeman & P. Edwards (Eds.), *Proceedings of the Ninth International Conference on Machine Learning*, pp. 363–368. San Mateo, CA: Morgan Kaufmann.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (in press) Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*.

Suddarth, S.C. & Kergosien, Y.L. (1991) Rule-injection Hints as a Means of Improving network Performance and Learning Time. In L.B. Almeida & C.J. Wellekens (Eds.), *Neural Networks/EURASIP Workshop 1990*, pp. 120–129. Berlin: Springer-Verlag.