Emergent Modularity and U-Shaped Learning in a Constructivist Neural Network Learning the English Past Tense

Gert Westermann (gert@cogsci.ed.ac.uk)
Centre for Cognitive Science
University of Edinburgh, 2 Buccleuch Place
Edinburgh EH8 9LW, Scotland UK

Abstract

A constructivist neural network model is presented that learns the past tense of English verbs. The model builds its architecture in response to the learning task in a way consistent with neurobiological and psychological evidence. The model outperforms existing connectionist and symbolic past tense models in terms of learning and generalization behavior, and it displays a U-shaped learning curve for many irregular verbs. The trained model develops a modular architecture with dissociations between regular and irregular verbs, and lesioning the different pathways leads to results comparable with neurological disorders. It is argued that the success of the model is due to its constructivist nature, and that the distinction between fixed-architecture and constructivist models is fundamental. Given this distinction, constructivist systems provide more realistic models of cognitive development.

Introduction

Models of learning the English past tense have in the past ten years become representative of different theories of language acquisition and cognition in general. While connectionist approaches (e.g., Rumelhart & McClelland, 1986; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993) have maintained that both regular and irregular past tense forms can be produced in a homogeneous architecture by a single process, dual-route accounts (Pinker, 1991) argue for two distinct mechanisms in different pathways, where regular forms are produced by a rule and irregular forms are stored in an associative memory. More recently, however, modular connectionist (Westermann & Goebel, 1995) and homogeneous symbolic (Ling & Marinov, 1993) models have been proposed.

While no working dual-route models exist (see Nakisa et al., 1997, for a theory of why such models do not work), most of the existing homogeneous models rely on a fixed, predefined architecture which is chosen specifically for the task at hand. However, as is argued below, such fixed-architecture systems are problematic both on neurobiological and learning theoretic grounds and might be limited in their usefulness as models of cognitive development. In this paper a constructivist neural network (CNN) model is presented that learns the English past tense by growing its architecture in response to the specific learning problem (see also Westermann, 1997). The performance of this network is evaluated against three existing models of past tense learning: the original network by Rumelhart & McClelland (1986, R&M), the improved backpropagation model by MacWhinney & Leinbach (1991, M&L), and the Symbolic Pattern Associator (SPA, Ling & Marinov, 1993), a symbolic decision-tree model. It is shown that the present constructivist model outperforms the existing symbolic and subsymbolic models both in terms of psychologically realistic learning and generalization behavior. Further, it is shown that the constructivist growth process leads to a modular architecture in which regular and irregular verb forms are dissociated and can be selectively impaired by lesioning different pathways of the network.

In the next section, the neurobiological and learning theoretic arguments for constructivist models of cognitive development are reviewed. Then, the network algorithm that was used in the simulations in this paper and the precise experimental setup are described. The Results section is concerned with the performance of the network in terms of learning, generalization, U-shaped learning, and emergent modularity leading to a dissociation between regular and irregular verbs. These results are then discussed in the final section.

Constructivist Cognitive Development

Cognitive development has recently been argued to closely correlate with the structural development of the cortex, with an increase in structural complexity leading to an increase in cognitive capacities (Quartz & Sejnowski, 1997). Moreover, the initial functional restrictions of a child's cognitive capacities seem to be essential for reaching adult competence (e.g., Elman, 1993). In order to understand the principles of cognitive development it is therefore important to take the mechanisms of brain development into account.

Recent work in this area has provided evidence that the development of cortex is activity dependent on different levels (see e.g., Fields & Nelson, 1992). Activity can determine the polarity of neurons, the rate and direction of dendritic and axonal growth, and the formation of synapses (e.g., Quartz & Sejnowski, 1997). Stabilization and loss of these synapses is also activity dependent (Fields & Nelson, 1992). It has also been shown that the cortex is not innately prespecified but readily adapts to process afferent signals from different domains (O'Leary, 1989). These results indicate that neural development proceeds in a *constructivist* way, with the neural organization of the brain being modified through constructive and regressive events by complex interactions between genetic predispositions and environmental inputs.

Cognitive development which is based on cortical development will thus proceed in the same constructivist manner, where activity dependent architectural modifications lead to increasingly complex cognitive representations.

Most significantly, research in learning theory (Quartz, 1993) has shown that incorporating activity dependent structural modification into a learning system is not just a way to tune performance, but leads to entirely different learning properties of that system, evading many of the problems that are associated with fixed-architecture systems. Any *a priori* choice of architecture severely limits the class of problems that can be learned by a model, which is often manifested in a trial-and-error approach to choosing the number of hidden units in a network for a specific task and which has led to the rejection of the concept of "learning" in favor of a "fixation of belief" on innate representations by some (e.g., Fodor, 1980). Constructivist systems, however, can overcome these limitations (Quartz, 1993).

Given the fact that the cortex develops in an activity-dependent way, and taking into account that systems developing in this way are fundamentally different from fixed-architecture systems, plausible cognitive models should likewise adapt their architecture in a way which is specific to the learning task. Such models can be called *constructivist*, reflecting their proximity to the constructivist developmental theories of Piaget in which structural modification of the learning system occurs in response to environmental input (see also Mareschal & Shultz, 1996).

The CNN which is presented in this paper models the acquisition of the English past tense in a constructivist process. The CNN is compared with other existing models, allowing for the empirical assessment of the suitability of constructivist models for the simulation of cognitive development.

The Constructivist Neural Network Model

For the cognitive simulations described here, a modified version of the constructivist Supervised Growing Neural Gas (SGNG) algorithm (Fritzke, 1994) was used because it incorporates constructive and regressive events which depend on the learning task, and because it provides mechanisms to produce outputs based on both the structure and on the identity of input signals, conforming to both neurobiological and psychological evidence.

The SGNG algorithm constructively builds the hidden layer of a radial basis function (RBF) network. RBF networks are different from backpropagation networks in that the hidden units have a Gaussian ('bell-shaped') rather than a sigmoid activation function. This allows each hidden unit to be active only for inputs within a certain range (as opposed to being active for all inputs above a certain threshold), thus forming a *receptive field (rf)* for a region of the input space. All input vectors are positioned in this space according to their feature values, and the hidden units are placed at different positions to cover the whole space. Hidden units will be activated by an input if it falls within their rf, and the closer the input is to the center of the rf the more the unit will be activated.

The problem in building RBF networks is to decide on the number and positions of the hidden units. The SGNG algorithm solves this problem by constructing the hidden layer during learning, adding units when and where they are needed.

The CNN starts with just two units in the hidden layer, each covering roughly half of the input space (see figure 1). The network tries to learn the task with this architecture (by adjusting the weights with e.g., gradient descent, the perceptron-rule, or quickprop), and when learning no longer improves performance, a new unit is inserted. The place where the new unit is inserted is determined by the classification error resulting from treating inputs within one rf as similar: the rf that previously caused the highest error is shrunk and the new unit is inserted next to it. The idea here is that a unit which produces a high output error is inadequate, and therefore more structural resources are needed in that area.

Figure 1 shows a hypothetical start and end state in a two-dimensional input space. While initially only two rfs cover the whole of the space, later hidden units have been inserted with different densities across the space to account for the specific learning task.

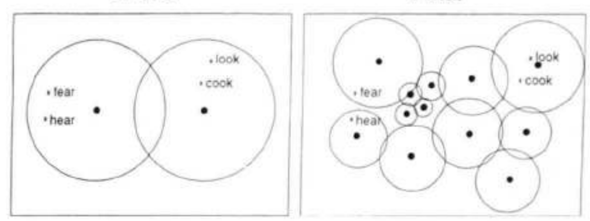

Figure 1: Receptive fields covering the input space at the beginning (left) and the end (right) of learning.

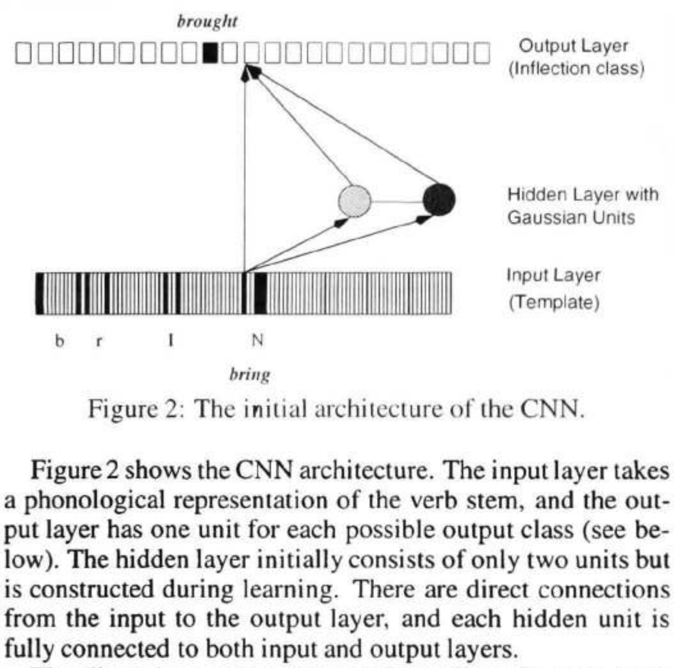

Figure 2: The initial architecture of the CNN.

Figure 2 shows the CNN architecture. The input layer takes a phonological representation of the verb stem, and the output layer has one unit for each possible output class (see below). The hidden layer initially consists of only two units but is constructed during learning. There are direct connections from the input to the output layer, and each hidden unit is fully connected to both input and output layers.

The direct input-output connections allow the past tense classes to be produced based on the structure of the input stem. By contrast, the (growing) hidden layer acts as a memory: it produces an output based on the identity and not the structure of the input. Initially, though, similar verbs will fall into the same rf even when they require different outputs (e.g., `hear` and `fear` requiring `heard` and `feared`, respectively). During the training of the network new rfs would be inserted in the area of such verbs, and eventually similar verbs with dissimilar past tense forms will be discriminated (see fig. 1). Where similar verbs have the same output class, however, (e.g., `look` and `cook` with `looked` and `cooked`), no new rfs will be inserted and a single rf will cover several verbs without producing an output error. Thus, the internal structure of the network will adapt to reflect the learning task. While taking advantage of a "starting-small" approach, this task-dependent adaptation is in contrast to other developing systems that incorporate a pre-programmed development such as an extension of short-term memory span, where the final architecture is independent from the specific problem being learned (see e.g., Elman, 1993).

Experiments

In order to allow for comparisons between different models, the present simulations employed the corpus from MacWhinney & Leinbach (1991), which was also used by Ling & Marinov (1993) in their SPA model. This corpus consists of 1,404 stem/past tense pairs of English verbs. The verbs were transcribed using UNIBET and, following MacWhinney & Leinbach (1991), represented in a template format containing slots for consonants and vowels. Table 1 shows examples of the phonological encoding of some verbs. Each phoneme was represented by ten features, such as *voiced*, *labial*, *dental* for consonants, and *front*, *center*, *high* for vowels. A template consisted of 18 slots, resulting in a 180-bit feature vector for the representation of each verb. For the output the verbs were classified according to how they form their past tense (adapted from Pinker & Prince, 1988). For example, the class $/x/ \rightarrow /U/$ contained the verbs `understand`, `withstand`, `overtake`, `stand`, `shake`, and `take`, and the class $/x/ \rightarrow /6/$ comprised `string`, `strike`, `swing`, `stick`, `fling`, `cling`, `spin`, `hang`, and `dig`. This classification resulted in 23 output classes, one for regular and 22 for irregular verbs. Each output unit of the network corresponded to one output class.

Table 1: Some examples for the template-encoding of verbs.

bring	`br-I-N------------`
explain	`---I-ksp--l--e-n--`
recognize	`r--E-k--I-gn-3-z--`
Template	`CCCVVCCCVVCCCVVCCC`

From the original corpus of 24,802 tokens, 8,000 tokens were randomly extracted according to the frequency of their past tense forms. The structure of the resulting corpus is summarized in table 2.

Table 2: The structure of the training corpus.

	Types		Tokens	
regular	943	(88.4%)	4579	(57.2%)
irregular	123	(11.6%)	3421	(42.8%)
total	1,066	(100%)	8,000	(100%)

Training was non-incremental: the whole training set of 8,000 stem/past-tense-class pairs was presented to the network in random order at each epoch. Hidden units were inserted depending on the learning progress (see previous section), and the network was tested for its performance on the training set prior to each insertion.

Results

This section describes the results of the learning experiment.

Learning

The network was trained for 4,055 epochs, at which point it correctly classified 99.6% (939 out of 943) of the regular verbs and 97.6% (120 out of 123) of the irregular verbs. Table 3 compares the learning results of the CNN with the R&M, M&L, and SPA models. While all models performed equally well for the regular verbs, the CNN outperformed the other models on the irregular verbs, closely followed by the symbolic SPA. This success seems to be a direct consequence of the ability of the CNN to allocate structure where needed, and thus specifically for the harder-to-learn irregular verbs. The precise structure of the trained CNN is analyzed below.

Table 3: Performance on training of the four compared models (extended from Ling & Marinov, 1993).

	R&M	M&L	SPA	CNN
Verb types	420	1,650	1,038	1,066
Percentage correct				
Total	97.0	99.3	99.2	99.3
Regulars	98.0	100.0	99.6	99.6
Irregulars	95.0	90.7	96.6	97.6

Generalization to Novel Verbs

The trained CNN was tested on its generalization to novel verbs. As Ling & Marinov (1993) have pointed out, testing the generalization ability of a model on existing verbs is misleading because irregular verbs are by their nature unpredictable. Therefore, in line with Ling & Marinov, the CNN was tested on a set of 60 pseudo-verbs which had also been tested on human subjects (Prasada & Pinker, 1993). These verbs consisted of blocks of ten which were prototypical, intermediate and distant with respect to existing regular and irregular verbs.

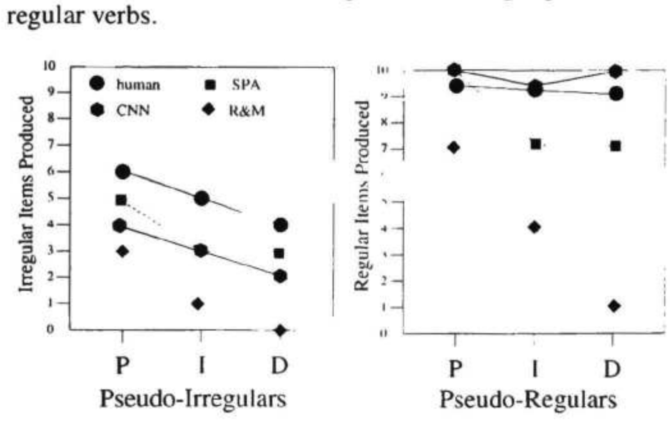

Figure 3: Generalization of the CNN to different classes of pseudo-verbs, in comparison with humans, the SPA, and R&M's network (extended from Ling & Marinov 1993). P = Prototypical, I = Intermediate, D = Distant.

The results of the generalization experiments are shown in figure 3. The CNN had a stronger tendency to regularize novel "irregular" pseudo-verbs than human subjects and performed similarly to the SPA. For "regular" pseudo-verbs the network performed very similar to human subjects and better than both the SPA and the R&M models: 29 of the 30 pseudo-regulars were regularized, and the one remaining verb, `pleem`, was equally strong classified as regular (`pleemed`) and belonging to the `sleep–slept` class, thus yielding `plemt`.

Developing Network Architecture

At the end of training the CNN had constructed a hidden layer consisting of 397 rf units, i.e., on average each rf accounted for 2.69 verbs. However, a closer analysis of the distribution of these rfs over the input space showed a large difference between regular and irregular verbs: the 123 irregular verb types were distributed over 107 rfs, i.e., each irregular verb claimed on average 87% of an rf (ranging from one to three verbs per field). By contrast, the 943 regular verb types were distributed over just 347 rfs (some of which covered both regular and irregular verbs), i.e., each regular verb claimed on average just 37% of an rf, with the number of regular verbs in a single rf ranging from one to twelve. This result clearly shows the advantage of constructivist as opposed to fixed-architecture models: a constructivist model can allocate its resources where they are needed, and there is no need for an *a priori* choice of the number of hidden units. Each hidden unit processes only a small subset of the verbs and can therefore learn the correct output efficiently. The final network architecture reflects the properties of the learning problem; in this case, a number of large rfs for the regular verbs and small, fine-grained rfs for the more difficult irregular verbs.

U-shaped Learning Curve

The most striking feature of past tense acquisition in children is the U-shaped learning curve: an unlearning of previously correct irregular past tense forms and their subsequent re-learning (e.g., `saw—seed/sawed—saw`), and a plausible model should follow this well-documented (Marcus et al., 1992) course. However, most of the existing models have been unable to provide a realistic account of the emergence of the U-shaped learning curve: whereas R&M relied on the assumption that the learning environment of a child changes from a first stage of mainly irregular verbs to a second stage of mainly regular verbs, and chose their training data accordingly, M&L's model could not account for any unlearning of irregular forms. In the SPA, U-shaped learning was achieved by the explicit manipulation of a learning parameter that controlled how many times a verb had to be seen in order to be memorized as an exception—if it occurred less often, it was overregularized. Besides "hard-wiring" the theory that children possess such a variable parameter, and using the resulting U-shaped learning curve as evidence for the same theory, this procedure also established a perhaps unrealistically direct relationship between the frequency of a verb and its overregularization. Plunkett & Marchman (1993) were able to show U-shaped learning in an environment where the training corpus was slowly expanded. However, it might be more plausible to assume that while the learning environment of the child is static, the child himself is undergoing changes that will influence the processing of the environmental input (see also Elman, 1993). Therefore, it was interesting to investigate the behavior of the CNN in a non-incremental training environment.

In fact the CNN displayed a U-shaped learning curve for many of the irregular verbs in the training corpus. A period of overregularization (i.e., a classification of the verb as belonging to the regular class) was preceded by a phase of correct classification; this was the case e.g., for `saw`, `sold`, `told`, `said`, `rang`, etc. At other points during training, many of these verbs were classified equally strongly as regular and as irregular, which corresponds to a regularization of the past tense form, e.g., `sawed` and `sanged`. The overall overregularization rate decreased from 25% very early on in training to 2.4% towards the end, although there were large differences between individual verbs. While no across-the-board U-shaped learning affected all irregular verbs simultaneously, corresponding to psycholinguistic evidence (Marcus et al., 1992) there was a phase of overregularizations at individual times for different verbs.

As with children, frequent verbs in the CNN were overregularized much less often than infrequent ones. The ten most frequent irregulars were overregularized on average only in 0.16% of all cases, whereas the ten least frequent ones had an average overregularization rate of 37.7%.

A further result corresponding to child language data concerned the protection from overregularization by similar sounding irregulars: the three verbs `hang`, `slide`, and `bear` were overregularized on average in 7.3% of all cases, whereas `ring`, `spring`, and `sing`, despite comparable token frequencies, were overregularized in only 0.9% of all cases.

The CNN model was thus successful in modeling the whole of the U-shaped learning curve including a correct production of past tense forms before their subsequent overregularization, and its performance corresponded to the details of children's past tense learning.

How does the U-shaped learning in the CNN occur? Since the verb set was held constant throughout training, the change in performance could only be a consequence of the internal reorganization of the network architecture. Initially, the network had only two hidden units, each of which roughly covered about half of all verbs with their varied past tense classes, and the CNN therefore had to rely on the direct input-output connections for producing the correct past tense classes. Given these restrictions the CNN initially learned the past tense classes of fewer irregulars but of most regular verbs.

During the training process, however, the CNN gradually grew its hidden layer, adding more receptive fields which lead to the reorganization of the internal representations (mainly of irregular verbs) from a structure-based (in the direct input-output connections) towards an identity-based (in the hidden layer) representation.

This construction process lead to a phase in which production of the irregular classes was partly taken over by the hidden layer, but the few receptive fields were large and included regular as well as irregular verbs, thereby causing errors even for irregular verbs that had initially been produced correctly through the direct input-output connections. This phase in the CNN, in which representations are relocated to insufficient resources, corresponds to the overregularization stage in children. It is evident that with this mechanism, different verbs would be overregularized at different times, depend-

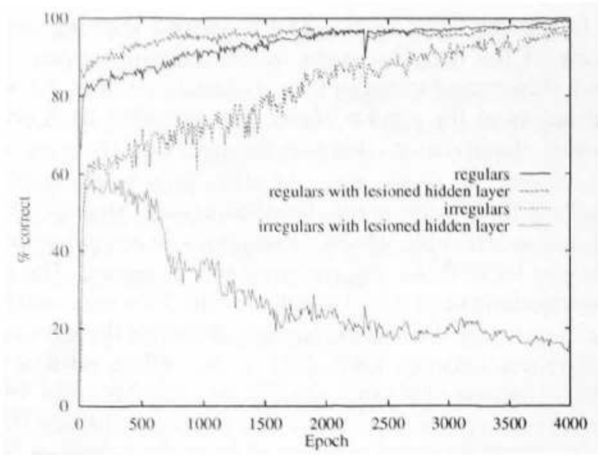

Figure 4: The learning curves for the regular and irregular past tense classes in the intact network and with a lesioned hidden layer.

ing on whether they had been allocated to an individual receptive field. The process of internal reorganization of the network's representations becomes evident in figure 4, which shows the learning curves for the regular and irregular past tense classes both in the intact CNN and after lesioning the hidden layer (by deleting the connections from the hidden to the output layer) when only the direct input-output connections were used.

Initially, with only a few hidden units, lesioning the hidden layer did not have a strong effect on network performance: with or without the hidden layer, initially about 60% of the irregular and 80% of the regular past tense forms were classified correctly. As the hidden layer grew, however, lesioning lead to a marked decrease in performance for the irregular verbs only, resulting in only 16% being classified correctly at the end of training when the hidden layer was lesioned. This result indicates that the representations of even initially correctly classified irregular verbs were shifted from the direct connections into the growing hidden layer, leading in many cases to the temporarily incorrect production of initially correct classes. The internal reorganization of the CNN due to a constructivist adaptation of its structure could thus account for the unlearning of initially correct outputs and therefore the U-shaped learning curve in the acquisition of the English past tense.

There was almost no effect of lesioning the hidden layer on the regular verbs, however, suggesting a developed dissociation between regular and irregular verbs. This dissociation is discussed in the next section.

Emergent Modularity

The functional dissociation between regular and irregular verbs that can be observed in psycholinguistic experiments but also in certain neurological disorders such as Specific Language Impairment (SLI) and Williams Syndrome, has led to the postulation of different pathways and subsystems for the production of these forms (e.g., Pinker, 1991). Pinker argued that irregular forms were produced in an associative memory whereas regular forms relied on the application of a mental rule. This rule was assumed to be independent from environmental input, possibly having a genetic basis.

However, models in which these two functionally distinct pathways have been hard-wired have been shown to work only with specific, unrealistic assumptions about the structure of a corpus with regular and irregular forms (Nakisa et al., 1997).

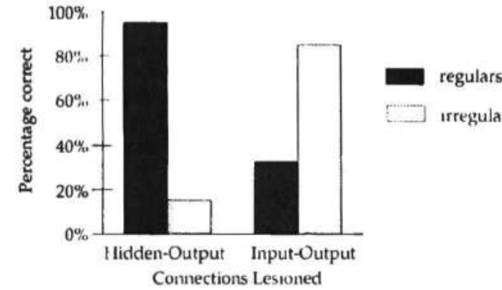

Figure 5: The effect of lesioning different pathways on the production of regular and irregular past tense classes.

In the CNN, a dissociation between regular and irregular verbs emerged as a direct outcome of the constructivist process. During learning, the two pathways, the direct input-output connections and the hidden layer, developed to take on specific functions. While the regular past tense class was produced through the direct connections, the irregular classes were produced mainly in the developing hidden layer. The double dissociation between regular and irregular verbs was demonstrated by lesioning both pathways selectively (figure 5): lesioning the hidden layer in the fully trained network left production of the regular verb class intact (97.45% correct) but severely impaired the production of irregular classes (only 16.26% correct). 90% of the irregular errors in this case were overregularizations, and in the other cases the network failed to produce any output class. By contrast, lesioning the direct input-output connections left production of the irregular classes only slightly impaired (83.74% correct) but had a marked effect on the regulars (only 32.13% correct). This functional modularity was not pre-specified, but developed solely through the construction of the hidden layer in response to the learning task and the resulting shift of the internal representations of the irregular verbs into this hidden layer.

These results might account for the language deficits observed in SLI and Williams Syndrome: SLI patients often have problems with converting present tense to past tense forms, with more pronounced difficulties for the regular forms. Pinker (1991) cited a study in which SLI children supplied 85% of irregular past tense forms but only 30% of regular forms correctly. In the CNN a very similar result is achieved by lesioning the input-output pathway. SLI could thus consist in a failure to develop or utilize this pathway.

Similarly, children suffering from Williams Syndrome seem to retrieve words in a deviant fashion and do not show the normal tendency to favor high-frequency words (Pinker, 1991). The syndrome can be accompanied by high overregularization rates, an effect which in the CNN can be explained with a partial or total lesioning of the receptive field hidden layer, or an insufficient development of this layer.

Discussion

The simulations described in this paper give empirical evidence that constructivist neural networks can model the acquisition of the English past tense more closely than other models which rely on fixed architectures. The ability of the CNN to develop its structure in response to the specifics of the learning task not only allowed it to allocate more structure to the difficult-to-learn irregular verbs, but also led to a U-shaped learning curve based on the internal reorganization of representations, and to an emergent functional modularity with dissociations between regular and irregular verbs. The model thus closely followed the developmental profile observed in children and reflected in its final architecture properties that can be found in adults through psycholinguistic experiments and in neurological disorders. Together with the theoretical arguments for constructivist learning these results offer compelling evidence for the usefulness of constructivist models in the study of cognitive development.

It is interesting to note that the previous model which came closest in its performance to the CNN is the symbolic SPA. However, this model is also constructivist: it dynamically builds a decision tree based on the training data. The results of the present simulations suggest that the fact that the SPA outperformed the R&M and M&L models was not primarily due to its symbolic architecture, but was based on its constructivist nature. Therefore, the dichotomy *fixed-architecture* vs. *constructivist* might be more fundamental than the traditional *symbolic* vs. *subsymbolic* distinction which previous past tense models have aimed to emphasize. Direct comparisons between symbolic and subsymbolic models can thus only be made either within or without the constructivist framework, with constructivist models conforming better to evidence from neural and cognitive development.

Although the SPA and the CNN performed similarly in terms of learning and generalization behavior, the CNN provided better explanations of U-shaped learning and showed an emergent modularity with dissociations between verb types. This might be due to the advantages of subsymbolic learning where low-level interactions between simple units lead to emergent complex behavior.

The results from both the present and previous simulations contradict the view that connectionist models entail a homogeneous architecture. The CNN develops a modular architecture with dissociations between verb types solely on the basis of a single learning mechanism. However, not all constructivist models will necessarily develop such a modular architecture. Therefore, a model of cognitive development should be classified along the three dimensions *symbolic–subsymbolic*, *modular–homogeneous*, and *fixed-architecture–constructivist*. Given this three-dimensional classification matrix, the present paper suggests that subsymbolic, modular constructivist systems provide the most realistic models of cognitive development in the child.

Acknowledgements

This research was supported by the ESRC (award no. R00429624342) and partly by the Gottlieb Daimler-und Karl Benz-Stiftung (grant no. 02.95.29)

References

Elman, J. L. (1993). Learning and Development in Neural Networks: The Importance of Starting Small. *Cognition*, *48*, 71–99.

Fields, R. D. & Nelson, P. G. (1992). Activity-dependent Development of the Vertebrate Nervous System. *International Review of Neurobiology*, *34*, 133–214.

Fodor, J. (1980). Fixation of Belief and Concept Acquisition. In M. Piattelli-Palmarini (Ed.), *On Language and Learning: The Debate between Jean Piaget and Noam Chomsky* (pp. 143–149). London and Henley: Routledge & Kegan Paul.

Fritzke, B. (1994). Fast Learning with Incremental RBF Networks. *Neural Processing Letters*, *1*, 2–5.

Ling, C. X. & Marinov, M. (1993). Answering the Connectionist Challenge: A Symbolic Model of Learning the Past Tenses of English Verbs. *Cognition*, *49*, 235–290.

MacWhinney, B. & Leinbach, J. (1991). Implementations are not Conceptualizations: Revising the Verb Learning Model. *Cognition*, *40*, 121–157.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in Language Acquisition. Monographs of the Society for Research in Child Development, Serial No. 228, Vol. 57, No. 4.

Mareschal, D. & Shultz, T. R. (1996). Generative Connectionist Networks and Constructivist Cognitive Development. *Cognitive Development*, *11*, 571–603.

Nakisa, R. C., Plunkett, K., & Hahn, U. (1997). A Cross-Linguistic Comparison of Single and Dual-Route Models of Inflectional Morphology. In P. Broeder & J. Murre (Eds.), *Cognitive Models of Language Acquisiton*. Cambridge, MA: MIT Press.

O'Leary, D. D. M. (1989). Do Cortical Areas Emerge from a Protocortex? *Trends in Neuroscience*, *12*, 400–406.

Pinker, S. (1991). Rules of Language. *Science*, *253*, 530–535.

Pinker, S. & Prince, A. (1988). On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Cognition*, *28*, 73–193.

Plunkett, K. & Marchman, V. (1993). From Rote Learning to System Building: Acquiring Verb Morphology in Children and Connectionist Nets. *Cognition*, *48*, 21–69.

Prasada, S. & Pinker, S. (1993). Generalization of Regular and Irregular Morphological Patterns. *Language and Cognitive Processes*, *8*(1), 1–56.

Quartz, S. R. (1993). Neural Networks, Nativism, and the Plausibility of Constructivism. *Cognition*, *48*, 223–242.

Quartz, S. R. & Sejnowski, T. J. (1997). The Neural Basis of Cognitive Development: A Constructivist Manifesto. *Behavioral and Brain Sciences*, *20*, 537–596.

Rumelhart, D. E. & McClelland, J. L. (1986). On Learning Past Tenses of English Verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing, Vol. 2* (pp. 216–271). Cambridge, MA: MIT Press.

Westermann, G. (1997). A Constructivist Neural Network Learns the Past Tense of English Verbs. In *Proceedings of the GALA '97 Conference on Language Acquisition*, (pp. 393–398)., Edinburgh, UK. HCRC.

Westermann, G. & Goebel, R. (1995). Connectionist Rules of Language. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, (pp. 236–241). Erlbaum.