Towards identifying principles for clinical intervention in developmental language disorders from a neurocomputational perspective

Anna Fedor (fedoranna@gmail.com)
Developmental Neurocognition Lab, Department of Psychological Sciences
Birkbeck College London, Malet Street, London WC1E 7HX, UK

Wendy Best (w.best@ucl.ac.uk)
Division of Psychology & Language Sciences, University College London,
Chandler House, 2 Wakefield Street, London WC1N 1PF, UK

Jackie Masterson (j.masterson@ioe.ac.uk)
Department of Psychology and Human Development, Institute of Education University of London
20 Bedford Way, London, WC1H 0AL, UK

Michael S. C. Thomas (m.thomas@bbk.ac.uk)
Developmental Neurocognition Lab, Department of Psychological Sciences
Birkbeck College London, Malet Street, London WC1E 7HX, UK

Abstract

We used a simple artificial neural network model, drawn from the domain of language development, to begin the work of understanding what principles underlie effective interventions for developmental disorders of language and cognition, from the perspective of neurocomputational mechanisms of development. The work aims to complement a clinical perspective of the principles of effective intervention. Our study explored the effectiveness of different types of intervention modeled as items added to the normal training set. We assessed whether best interventions were specific to problem domains, specific to deficit types, and/or dependent on when in development they take place. While the model was highly simplified, it represents a first step in seeking to understand how atypical internal representations may be reshaped by alternative training regimes. The next step is to scale up the simulations to more realistic models of specific task domains within language acquisition.

Keywords: Artificial neural networks, developmental language disorders, intervention.

Introduction

This paper represents the beginning of a project whose ultimate aim is to establish general principles for guiding intervention in developmental disorders of language and cognition. Our belief is that an understanding of neurocomputational mechanisms of learning and development can contribute to the establishment of such principles. Our initial work focuses on developmental disorders of language. Law et al. (2007) noted that in the practice of speech and language therapy, theories of the causes of deficit play a relatively minor part in guiding interventions, and the interventions employed are diverse. Where interventions are successful, Lindsay et al. (2011) commented that there have been no studies attempting to distil the active ingredients of intervention, presumably because positive outcomes are the main focus.

Computational models of development, particularly those employing artificial neural networks (ANN), have provided hypotheses about the mechanistic bases of language deficits. For example, Harm, McCandliss and Seidenberg (2003) demonstrated how limited connectivity in the phonology component of a reading model produced a system with symptoms of dyslexia. In a model of inflectional morphology, Thomas (2005) demonstrated how shallow sigmoid activation functions yielded processing units that were insensitive to small changes in the input and networks exhibiting developmental delay. Progress of this type motivated Daniloff (2002, p.viii) to comment ‘ANN theory will … form the backbone of much of language therapy in the near future’. However, research and practice have yet to repay this optimism (though see Poll, 2011, for renewed attempts to make these links). Only one computational study has systematically explored the efficacy of a single intervention (in Harm et al.’s 2003 reading model). Another non-developmental study sought to show how an adult model of aphasia could guide actual interventions depending on patients’ error patterns (Abel et al., 2007).

A computational approach has generated a growing understanding of environmental factors that influence learning in typical development (Borovsky & Elman, 2006; Gomez, 2005; Onnis et al., 2005), including the importance of factors such as the frequency of training items, their variability, and the provision of novelty in familiar contexts. However, there has yet to be a consideration of how these factors interact with learning systems containing the sorts of atypical computational constraints that lead to impoverished internal representations, and in turn, behavioral deficits compared to typically developing children. It is yet a further step to link such an understanding with the diverse activities that tend to be used by clinicians in speech and language therapy, including such activities as modeling, forced alternatives, repetition, visual approaches to support oral language, and reducing distractions (Law et al., 2011).

We therefore believe that the time is right to being the work of constructing a foundation of neurocomputational understanding of how altered training environments can ameliorate the effects of atypical mental representations. It should pursue the following road map: the construction of models of typical development of language abilities; the creation of various atypical versions that exhibit different error patterns, based on the compromised computational constraints; an investigation of the kinds and timings of training interventions; and use of these insights to predict principles of effective intervention. Simultaneously, it is important to crystallize insights from clinicians about what interventions they find effective under different circumstances, then to make links between these two types of principle.

We begin the former strategy here, via the study of a very simple learning system, drawn from the domain of language development, and chosen because it provides insight into the emergence of internal representational states. We exposed this learning system to two different problem domains. We then applied two different deficits to the start state of the system, creating a 2x2 design of problem domain and deficit type. We next attempted a range of interventions, by altering the training environment at different stages of development. We then asked four questions: (1) Which types of intervention are best? (2) Does the timing of intervention matter? (3) Does the best intervention method depend on the type of underlying deficit? And, (4) Does the best intervention method depend on the nature of the problem domain?

Methods

For our target system, we chose a model employed to study the emergence of minority defaults in inflectional morphology (Forrester & Plunkett, 1994). The architecture is shown in Figure 1. There were two input units, generating a two-dimensional input space in which categories had to be learnt. We considered the simple case where the model had to learn three categories, represented by three localist output units. The network had fifty hidden units, allowing for high-dimensionality of its internal representations.

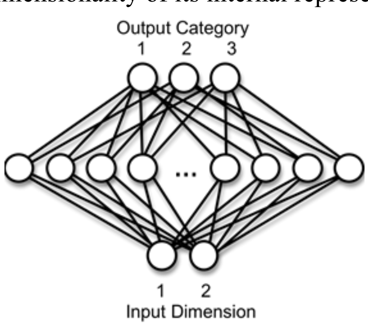
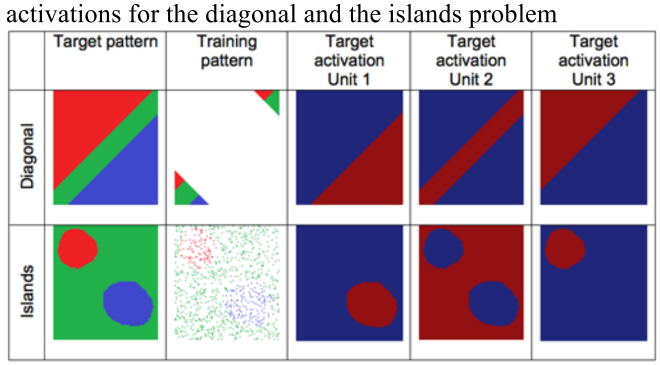


Figure 1. Network architecture (not all hidden units shown)

We investigated a range of categorization problems, eventually settling on one that emphasized regularity and one that emphasized idiosyncrasy. In the regular problem, networks had to learn a diagonal band of the input space as one category, and the regions either side as two others. In the idiosyncratic problem, the network had to learn that two 'islands' of the input space represented two categories, and the region around them a third. The input space was divided into 100 units, varying between -0.5 and +0.5, on each input unit, specifying 10,000 points in input space. Training on each problem corresponded to about 10% of the possible items. For the diagonal problem, these were sampled from either end of the diagonal. Testing was on the full set, so that the network was required to interpolate between the ends of the diagonal. For the island problem, patterns were sampled at random across the input space. The problems and training sets are illustrated in Table 1.

Table 1. Target patterns, training patterns and target activations for the diagonal and the islands problem



Typical development (TD): Output units were considered active if their output was greater than 0.6 and inactive if their output was smaller than 0.4. An item was scored correct if all three output units had the correct response. Networks were trained using the backpropagation algorithm (learning rate=0.1, momentum=0.3) to a performance criteria of 100% correct in case of the diagonal and 97% in case of the islands, which was reached at 843 epochs (SD=254) and 733 epochs (SD=84), respectively. Across networks with different random seeds, we identified different qualitative phases of development, whose timing could vary to some extent. The four phases are illustrated in Figure 2, along with the emergence of the internal representations supporting this behavior.

Deficits: We initially explored deficits in connectivity, shallow sigmoid, processing noise, learning rate, and number of hidden units, to produce development disorders. We selected the first two for further investigation, as representing cases of marked atypicality (connectivity) versus delay (sigmoid), and corresponding to deficits used by Harm et al. (2003) and Thomas (2005) in models of reading and inflectional morphology, respectively. In the connectivity manipulation, initial network connectivity was set at 30% instead of 100%. In the sigmoid manipulation, the temperature of the activation function on hidden and output units was set to 0.5 instead of 1 (see Thomas, 2005).

Intervention: We designed interventions as items to be added to the training set, on the assumption that intervention complements rather than replaces normal experience.

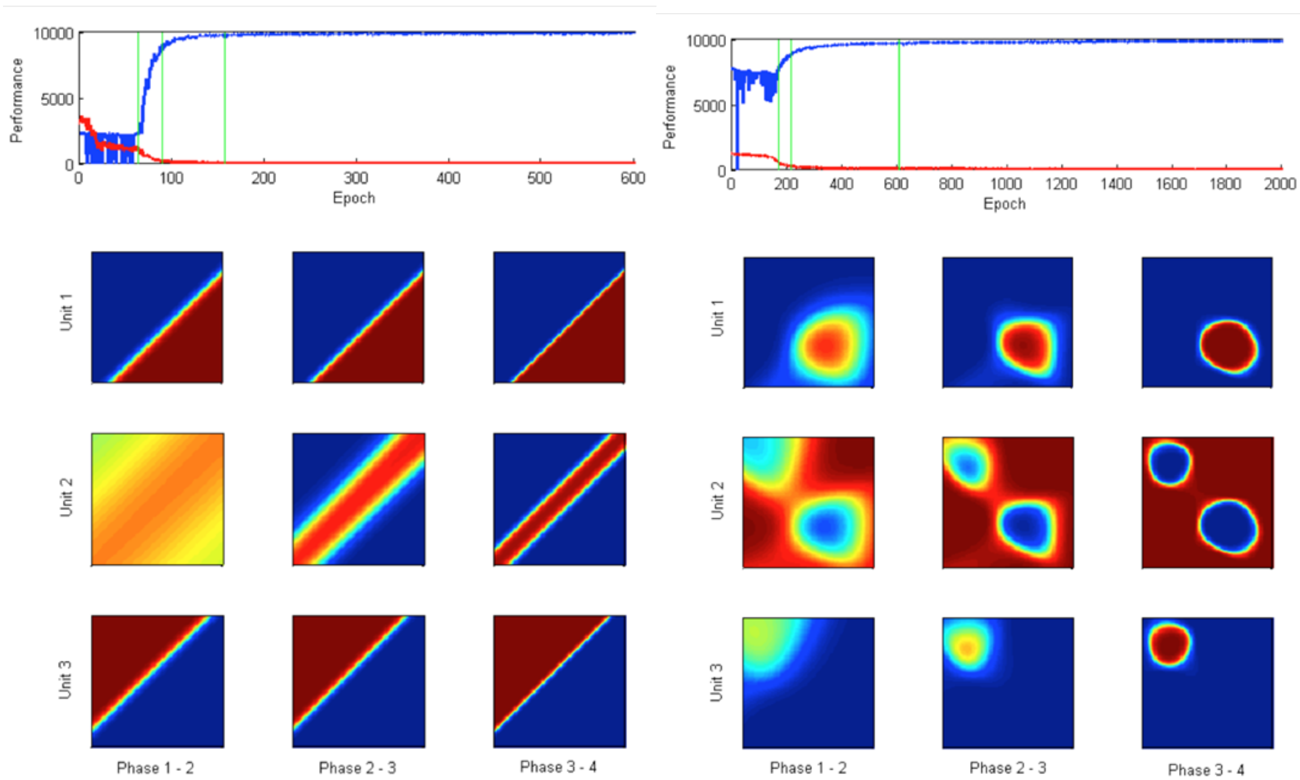


Figure 2. Developmental trajectories and phases for learning the diagonal (left) and the islands (right). Top figure: performance (blue) and mean square error (red) across development. Phase boundaries are indicated by green vertical lines. Second to fourth row of figures: snapshots of activation patterns of Output unit 1 to 3, respectively, at phase boundaries. Activation values are color-coded as temperature plots: red and blue indicates activation close to one and zero, respectively.

Interventions were designed to add sampling across the input space, to add training in areas that were 'prototypical' or central to each category, or in areas that demarcated category boundaries. Interventions are illustrated as part of Table 2. Interventions added 7% of items in the diagonal task and 10% of items in the islands task.

Intervention 1 contained the same type of information as the original training set in both tasks: a bigger "corner" for the diagonal and more randomly distributed items for the islands. Intervention 2 provided a transect of all three categories along the left diagonal. Intervention 3 and 4 contained four patches, each containing items from just one of the categories; they were further apart in Intervention 3 and closer together in Intervention 4. Intervention 5 contained items only from the boundaries of the categories. Intervention 6 contained random items for the diagonal and items from between the islands for the islands. We chose the latter because the area between the islands tended to be difficult to learn. Interventions were applied to the atypical model at different phases, either in the middle of phase 1, 2, or 3. Interventions at phase 4 were generally ineffective according to our previous experiments so they were not tested here. The six interventions were applied separately to each atypical network (N=10 replications). Training continued for 2000 epochs or until performance reached the criterion.

Results

We calculated the difference between the performance of the models in all intervention conditions and the corresponding network at the same point in training without intervention, to generate an improvement score. We refer to the diagonal problem with the connectivity deficit as scenario 1, the islands problem with connectivity deficit as scenario 2, the diagonal problem with low temperature as scenario 3, and the islands problem with low temperature as scenario 4. All conditions in all scenarios passed the Kolmogorov-Smirnov normality test. We used one sample t-tests with Bonferroni correction to see whether the mean improvement in intervention conditions differed from zero. Since there were 6 interventions x 3 phases = 18 comparisons in each scenario, we accepted t-tests as being significant only if they had a p-value lower than 0.0028 (0.05/18 = 0.00277). We then used repeated measures ANOVA (within-subjects factors were 'phase' and 'intervention'; there were no between-subjects factors or covariates) to assess whether the timing or the type of intervention had an effect on improvement. Mean improvement and its confidence interval in each condition are shown in Figure 4. Since there was considerable individual variability, we also looked at the networks one-by-one and counted the number of networks where there was positive improvement (i.e., the change was larger than zero). A summary of these results is provided in Table 2.

Low temperature models represent a milder deficit corresponding to delay, where intervention increased the speed of development. The final performance did not improve markedly, because it was already quite high. Low connectivity was a more serious problem, and intervention was not successful in most cases. Even if there was an improvement, performance did not get close to that of the TD models, especially in the case of the more difficult islands task. None of the interventions caused significant improvement in the islands task with this deficit; in the diagonal task, Intervention 2 was successful in phase 3 (see an example in Figure 3) and Intervention 6 was successful in phase 2 and 3. This said, intervention was still effective in *individuals* even when the improvement in group performance was not significant. Many networks improved due to intervention, illustrating heterogeneity in response to intervention even in highly simplified learning systems.

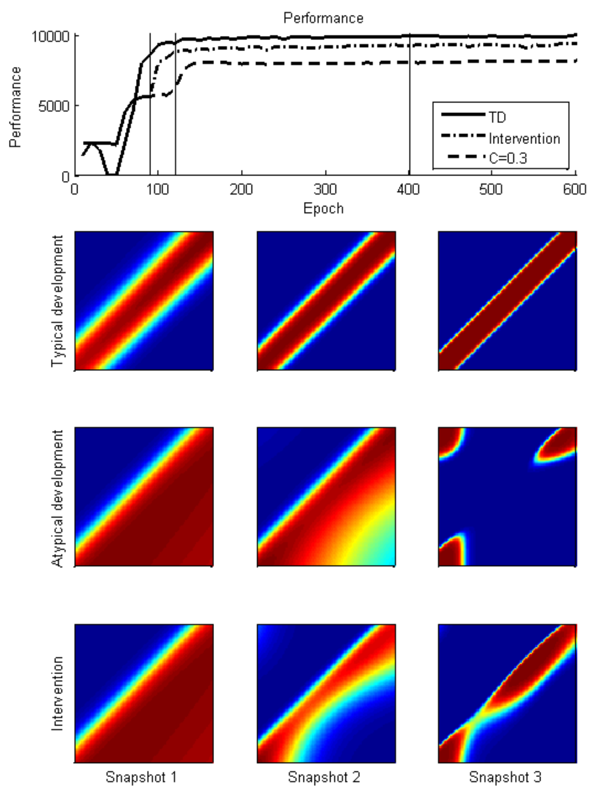


Figure 3. Developmental trajectories and internal representations in a typical case, an atypical case with low connectivity and the same atypical case with intervention. Top figure: Developmental trajectories; intervention commenced where the Intervention and the C=0.3 lines part. Vertical lines show epochs at which snapshots were taken. Colored figures: snapshots of the activation pattern of Unit 2 in the three cases.

With respect to individual interventions, the results partly overlapped across deficits. Adding more items of the same type or a transect of the categories (Intervention 1 and 2) were the best choices for improving performance, while providing separate patches of the categories was the worst (Intervention 3 and 4). Providing items from around the boundaries of the categories (Intervention 5) improved performance only in the low temperature deficit, not the reduced connectivity deficit, but now for both problems, suggesting a deficit-specific intervention. This makes sense, since it served to "sharpen" category boundaries, which was the main weakness of the low temperature condition. By contrast, we can think about Intervention 6 for the island pattern (items from between the islands) and Intervention 1 for the diagonal pattern (corners of the diagonal) as task-specific treatments. With respect to timing, Figure 4 shows that some interventions for some deficits were more effective at earlier phases, but this was not uniform. However, our pilot experiments showed that in phase 4 (the 'adult' state), most interventions had no effect.

In sum, the severity of the deficit and the difficulty of the task both influenced the outcome of interventions. The same interventions were among the best across deficits and tasks (random items, items of the same type and items from the transect) and the same interventions were the worst (separate patches), even though they improved the milder deficit in some conditions. We also found deficit-specific (items from the boundaries of the categories) and task-specific interventions (bigger corners for the diagonal).

Discussion

In our investigation of intervention, we addressed four questions. First, which types of intervention were best? The answer was, those that sampled the whole problem space, or those that provided a representative slice across all categories. Second, did the timing of intervention matter? The answer was yes and no: yes to the extent that our interventions were mostly ineffective in the 'adult' or fully trained networks; no in that timing effects that depended on the earlier phases of development were observed in only 5 of 24 problem-deficit-intervention combinations. Third, did the best intervention method depend on the type of underlying deficit? We did identify an intervention that worked for one deficit type and not the other across the two problems (Intervention 5). Finally, did the best intervention method depend on the nature of the problem domain? We identified another intervention that worked for one of the problem domains across the two deficits (Intervention 6).

In reality, behavioral interventions to remediate developmental disorders of language and cognition can be multi-faceted. They are usually interactional and social, and involve emotional and motivational factors in the child, as well as cognitive factors. There are myriad causes of variability in children's abilities, be they biological, psychological, environmental, or social – factors that must be considered in planning preventions or interventions (Beauchaine, Neuhaus, Brenner & Gatzke-Kopp, 2008). Clinical practice is driven by a range of principles including the emerging evidence base and the therapeutic setting, as well as the child and family's goals.

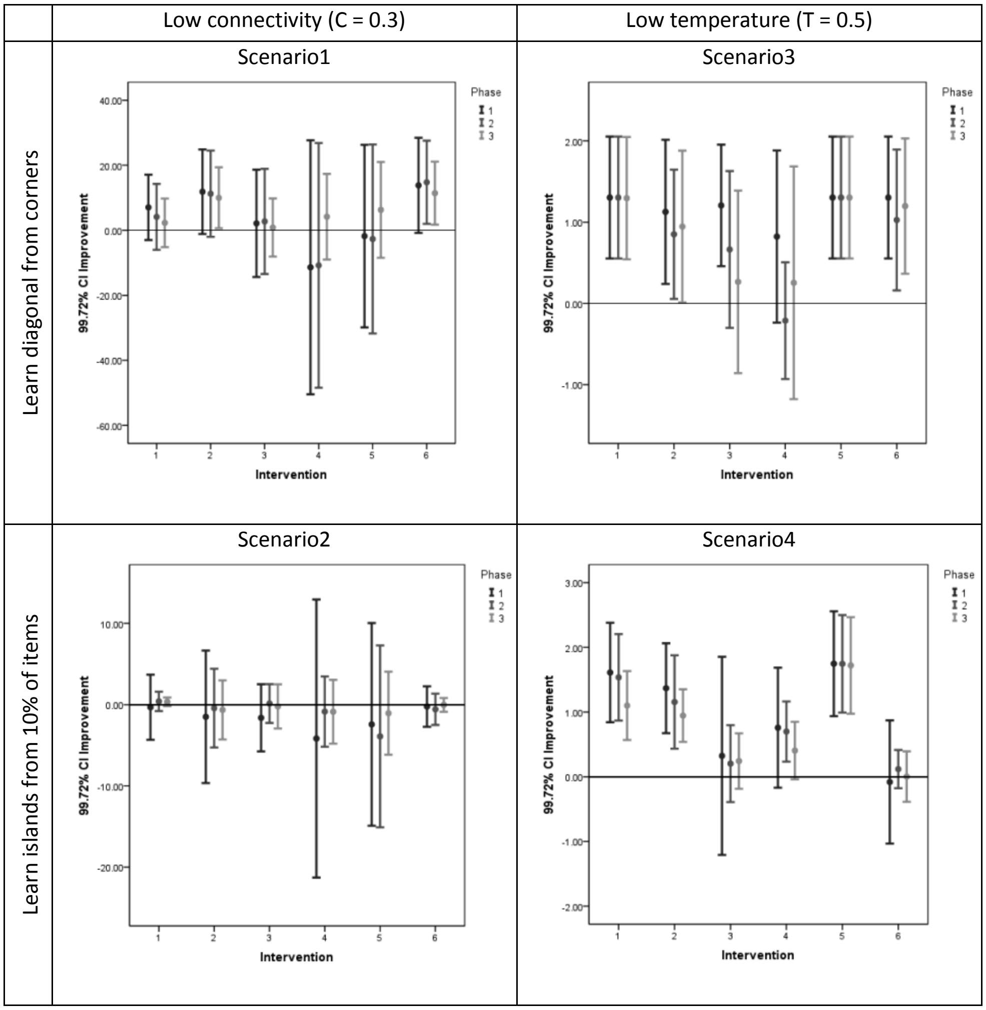


Figure 4. Mean improvement due to intervention in all scenarios, 18 conditions each (3 phases x 6 intervention types). Bars represent 99.72% confidence interval of the mean; t-tests are significant ($p < 0.0028$) if the bars do not touch the zero line.

Table 2. Summary of the results of t-tests and examination of individual performance. In each cell of the matrix, numbers in the top row in bold represent phases in which a particular intervention was successful according to the t-test. The bottom row of numbers lists the phases in which more than 7 networks improved due to intervention.

Task / Scenario	**Target pattern**	**Intervention** 1	2	3	4	5	6
Diagonal							
Islands							
Scenario 1	Diagonal C = 0.3	- 1,2	**3** 1,2,3	- -	- -	- 3	**2, 3** 1,2,3
Scenario 2	Islands C = 0.3	- 1,2,3	- 2	- 2	- -	- -	- -
Scenario 3	Diagonal T = 0.5	**1,2,3** 1,2,3	**1,2** 1,2,3	**1,2** 1	- -	**1,2,3** 1,2,3	**1,2,3** 1,2,3
Scenario 4	Islands T = 0.5	**1,2,3** 1,2,3	**1,2,3** 1,2,3	- 3	**2** 1,2,3	**1,2,3** 1,2,3	- -

Within approaches targeting speech and language needs directly, the clinician may form a hypothesis as to (i) the nature of the difficulty and (ii) what will be optimally effective for a child. The results of intervention will further refine these hypotheses. Nevertheless, the quality of neurocomputational mechanisms of learning and development is a key constraining factor, given that these mechanisms underlie behavior, and given that their plasticity is crucial in achieving remediation

Our model was, of course, highly simplified. It employed trivial learning domains and a single mechanism. Real neurocognitive systems have multiple interacting components operating on complex representations. Developmental limitations to different sub-sets of components may produce different deficits, and require behavioral interventions that target individual components.

Moreover, there are other complexities to consider, such as the possibility of compensatory pathways, the facilitatory role of supporting context, and the role of feedback. Interventions were modeled as items added to the training set, on the assumption that intervention complements rather than replaces normal experience. But, one could argue that other interventions manipulate computational properties more directly, through motivation, attention, or pharmacological means. These are certainly avenues to consider in future modeling work.

Despite the limitations, the current work was legitimated on two grounds. First, we couldn't anticipate in advance the answers to our four questions, even given the simplicity of the model learning system, suggesting a lack of even basic knowledge of how intervention may reshape mental representations that have developed under atypical computational constraints. Second, there is an emergent literature investigating the principles that guide clinical intervention (Fey et al., 2003; Law et al., 2007). Modeling can spur the elucidation of such principles by aiding our theoretical understanding of the key issues, the form the principles will likely take, and possible limits on their scope.

Acknowledgments

This work was supported by ESRC grant RES-062-23-2721.

References

Abel, S., Willmes, K. & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology*, *21(5)*, 411-447.

Beauchaine, T. P., Neuhaus, E., Brenner, S. L., & Gatzke-Kopp, L. (2008). Ten good reasons to consider biological processes in prevention and intervention research. *Development and Psychopathology*, *20*, 745-774.

Borovsky, A. & Elman, J. L. (2006). Language input and semantic categories: a relation between cognition and early word learning. *Journal of Child Language*, *33(4)*, 759-790.

Daniloff, R. G. (2002). *Connectionist approaches to clinical problems in speech and language*. Erlbaum: Mahwah, NJ

Fey, M. E., Long, S. H., & Finestack, L. H. (2003). Ten principles of grammar facilitation for children with specific language impairments. *American Journal of Speech and Language Pathology*, *12(1)*, 3-15.

Forrester, N. & Plunkett, K. (1994). Learning the Arabic plural: the case for minority default mappings in connectionist networks. In A. Ramand & K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 319-323). Mahwah, NJ: Erlbaum.

Gomez, R. L. (2005), Dynamically guided learning. In M. Johnson & Y. Munakata (Eds.) *Attention and Performance XXI* (pp. 87—110). Oxford: OUP.

Harm, M. W., McCandliss, B. D. & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, *7*, 155-182.

Law, J., Campbell, C., Roulstone, S., Adams, C. & Boyle, J. (2007). Mapping practice onto theory: The speech and language practitioner's construction of receptive language impairment. *International Journal of Language and Communication Disorders*, *43*, 245–63.

Lindsay, G., Dockrell, J., Law, J., & Roulstone, S. (2011). *Better communication research programme: 1st interim report*. Technical report, UK Department for Education.

Onnis, L., Monaghan, P., Christiansen, M., & Chater, N. (2005). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In: K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1047-1052). Mahwah, NJ: Erlbaum.

Poll, G. H. (2011). Increasing the odds: Applying emergentist theory in language intervention. *Language, Speech, and Hearing Services in Schools*, *42*, 580-591.

Thomas, M. S. C. (2005). Characterising compensation. *Cortex*, *41(3)*, 434-442.