DOCUMENT RESUME

ED 338 051 FL 019 712

AUTHOR	MacWhinney, Brian; Leinbach, Jared
TITLE	Implementations Are Not Conceptualizations: Revising the Verb Learning Model.
PUB DATE	Jul 90
NOTE	11p.; In: Papers and Reports on Child Language Development, Number Twenty-nine. California, Stanford University, 1990. p86-95.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Child Language; *Concept Formation; English; *Language Acquisition; Language Research; *Linguistic Theory; Models; *Tenses (Grammar); *Verbs

ABSTRACT

A model of the child's learning of the past tense forms of English verbs is discussed. This connectionist model takes as input a present-tense verb and provides as output a past tense form. A new simulation is applied to 13 problems raised by critics of the model, presented as fundamental flaws in the conceptualizations underlying connectionism. The new simulation uses a new input representation based on feature/slot units and two views on input words, a new architecture using identity mappings and hidden units, a new learning algorithm, and an input corpus that includes all five cells of the English verb paradigm. Together the changes led to a vast improvement in the model's performance. All problems but those dealing with the issue of direct access within connectionist nets were addressed successfully. It is concluded that the connectionist models are extremely useful ways of characterizing the learning of inflectional systems, and that the critiques erred by confusing conceptualizations of the model with implementations. (MSE)

Implementations are not Conceptualizations: Revising the Verb Learning Model

Brian MacWhinney
Jared Leinbach
Carnegie Mellon University

A recent issue of Cognition was devoted to an in-depth criticism of the connectionist agenda. At the focus of much of this criticism was a PDP simulation by Rumelhart and McClelland (1986) of the acquisition of the phonological form of the English past tense. A lengthy article by Pinker and Prince (1988) was devoted entirely to the critique of the R&M verb learning model. A somewhat shorter article by Lachter and Bever (1988) devoted most of its pages to the R&M verb learning model, while also critiquing four other connectionist models in somewhat less detail. The chapter that served as the target of this extensive criticism had as its goal the formulation of an computationally explicit connectionist model of the child's learning of the past tense forms of English verbs. This "verb-learning" model of Rumelhart and McClelland (henceforth R&M) was capable of taking as input a present tense verb, such as "ring," and providing as output a past tense form, such as "rang." It did this without any overt encoding of a set of rules and without any formal construction of morphological paradigms.

The arguments against this model presented by Pinker and Prince (henceforth P&P) and Lachter and Bever (henceforth L&B) were well-constructed and thoughtfully developed, and their critique has generated a great deal of useful discussion. Moreover, there is good reason to agree with many of the detailed aspects of the two critiques. However, these authors are mistaken in thinking that their criticisms call into question the general connectionist conceptualization, rather than merely casting doubt on a particular connectionist implementation. In this regard, it is crucial to note that P&P believed that their analysis was quite profound (p. 82).

> "We will conclude that the claim that parallel distributed processing networks can eliminate the need for rules and for rule induction mechanisms in the explanation of human language is unwarranted. In particular, we argue that the shortcomings are in many cases due to central features of connectionist ideology and irremediable; or if remediable, only by copying tenets of the maligned symbolic theory. The implications for the promise of connectionism in explicating language are, we think, profound."

L&B also believed that they had isolated a fundamental weakness in all connectionist approaches to language learning (p. 243).

> "The connectionist models we have considered arrive at rule-like regularities in language behavior only insofar as the models already contain architectures and devices explained in humans by mental representations of categorical rules."

In this paper, we will consider whether these analyses are as profound as their authors claimed them to be.

1. The Criticisms

P&P and L&B argued that claims made by R&M regarding the success of the verb learning model are extremely misleading. They expressed their respective cases against the verb learning model in terms of a set of points. P&P (p. 81) presented at least ten specific problems which they believed were incorrectly addressed by the R&M model. These problems are repeated below in quotation marks. Numbers and problem names have been added for ease of reference.

1. **The u-shaped learning problem:** "Rumelhart and McClelland's actual explanation of children's stages of regularization of the past tense morpheme is demonstrably incorrect."
2. **The "ated" problem:** "Their explanation for one striking type of childhood speech error is also incorrect."
3. **The "hit-hit" problem:** "Their other apparent successes in accounting for developmental phenomena either have nothing to do with the model's parallel distributed processing architecture, and can easily be duplicated by symbolic models, or involve major confounds and hence do not provide clear support for the model."
4. **The "algalgal" problem:** "The model is incapable of representing certain kinds of words."
5. **The "slit-silt" problem:** "It is incapable of explaining patterns of psychological similarity among words."
6. **The "brag-grab" problem:** "It easily models many kinds of rules that are not found in any human language.
7. **The phonological regularities problem:** "It fails to capture central generalizations about English sound patterns. It makes false predictions about derivational morphology, compounding and novel words."
8. **The homophony problem:** "It cannot handle the elementary problem of homophony."
9. **The convergence problem:** "It makes errors in computing the past tense forms of a large percentage of the words it is tested on."
10. **The regular pattern problem:** "It makes incorrect predictions about the reality of the distinction between regular rules and exceptions in children and in languages."

Problems 1, 4, 5, 7, and 10 were also raised by L&B. In addition, L&B provided a second form of criticism that is quite different from that developed by P&P. L&B claimed that the R&M verb learning model achieved much of its success by using a variety of TRICS (The Representations It Crucially Supposes). They believed that, together, these TRICS led to a cryptoembodiment of rules with the connectionist net. These TRICS all relate to the design and interpretation of the Wickelfeatures in the verb learning model. L&B complained that the selection of 460 particular Wickelfeatures from a possible set of about 2000 involved a variety of decisions that tended to reconstitute traditional segmental phonemic information. L&B used this observation to argue that connectionist architectures necessarily contain cryptoembodiments of rules.

11. **The cryptorule problem**: The selective development of an input representation can lead to the cryptoembodiment of rules in PDP nets.

There are two additional problems with the R&M verb learning model that were not raised by either P&P and L&B.

12. **The early noise problem**: In the R&M model, many epochs of training were required before the model would output phonological forms close to those of recognizable words. There is indeed an early period in child phonology when words have a very indistinct shape. However, when children are working out the various forms of the verb, most of these phonological limitations have been overcome. To be sure, early past tenses include many overregularizations and mismarkings, but these errors are typically modifications of the basic form of the verb, rather than phonologically inarticulate forms.

13. **The direct access problem:** The R&M model works by converting one phonological representation into another phonological representation. This mode of access takes a "basic" form and uses it to find a "derived" form. This is certainly one of the ways in which we can access the correct past tense of a verb. However, we can also access words directly through meaning. It is not yet clear how connectionist models can simulate direct access in a theoretically interesting way.

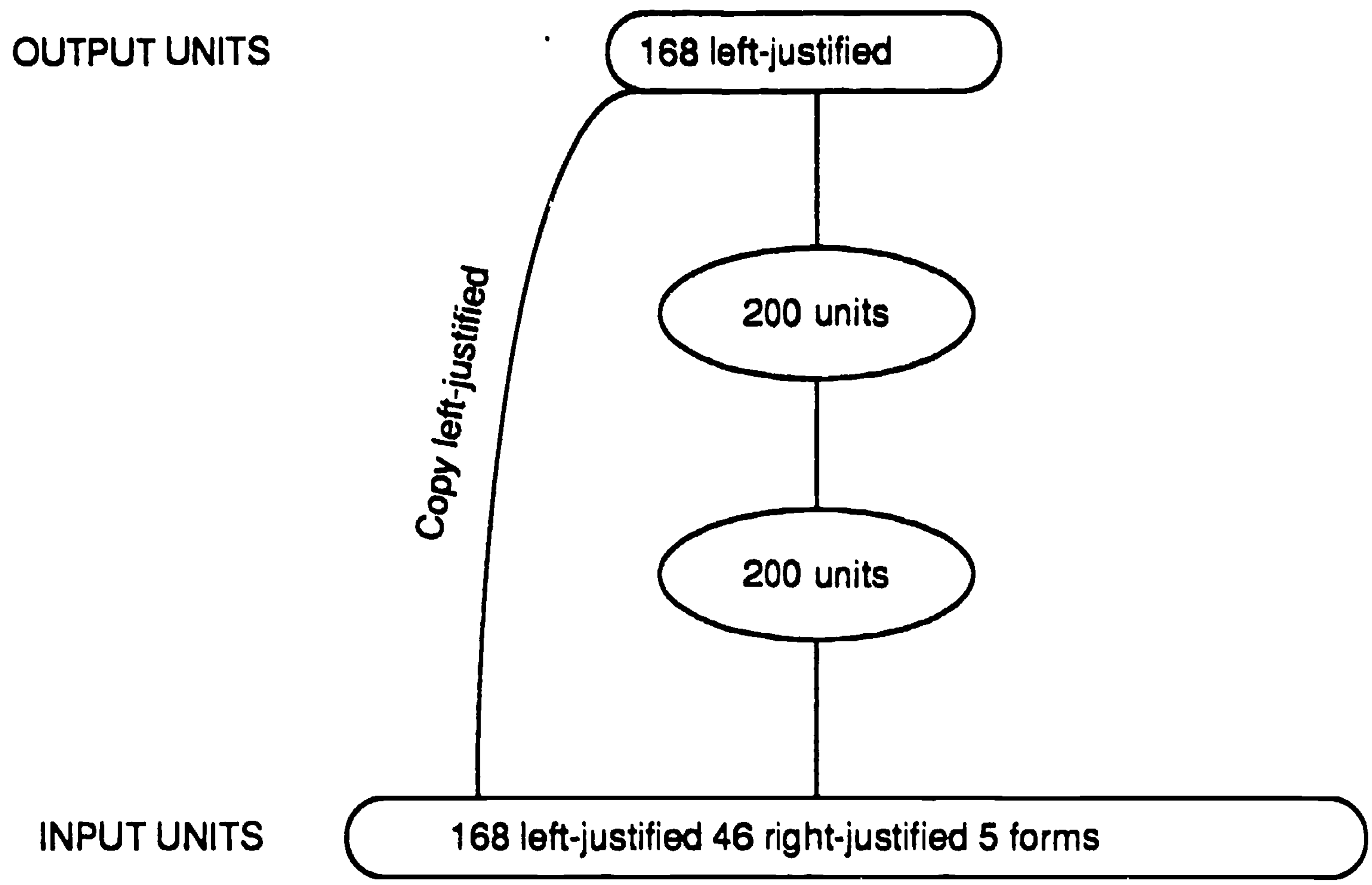
2. The new simulation

The new simulation uses a new input representation, a new output representation, a new network architecture, a new learning algorithm, and a new input corpus. Each word is represented by a trisyllabic left-justified pattern and a monosyllabic coda in a right-justified pattern. The trisyllabic pattern takes the form CCCVVCCCVVCCCVVCCC, where C stands for consonant and V stands for vowel. The coda pattern takes the form VVCCC. The way in which sounds are filled into these slots is explained below. Vowel nuclei are composed of up to two segments and consonantal clusters are composed of up to three segments. If a particular segment is not actually present in a word, its features are simply left off.

The 14 vowels of English are represented by eight distinctive features. Since there are six possible vowel slots, there are a total of 48 units dedicated to vowels. The 22 consonants of English are represented by ten distinctive features. Since there are twelve possible consonantal slots, there are a total of 120 units dedicated to consonants. Together, the 120 consonantal units and the 48 vocalic units yield a combination of 168 feature/slot units for the left-justified representation and another 46 feature/slot units for the right-justified coda for each input word. The input also includes five units dedicated to the five cells in the paradigm for English verbs. There is one unit each for the present, the past, the past participle, the present participle, and the third person singular present.

The output of the network is simply the 168 feature/slot units of the left-justified phonological form. Between the input and output units there are two pools of 200 "hidden" units. These units are called "hidden" because they have no direct interpretation in either input or output terms. Between adjacent pools, every unit of one pool is connected to every unit of the other pool. The model uses two pools of hidden units, because a model which had only one pool of units did not do as well at learning the training set.

The final feature of the network is a set of identity-mapping connections between the left-justified input and the output. The network was designed to treat the learning of the "derived" forms of the verb as modifications of the phonological form of the "basic" present tense. The idea is that the child assumes that the past tense is somehow a modification of the present. This is done by including a set of connections that "copy" the left-justified phonological form of the input directly onto the output. This copying only sets a weak bias on the activation of the output units. This bias can be overcome with learning. Indeed, as we will see, the initial bias is usually overcome within a few trials.



2.4. The learning procedure.

A major difference between the current model and the R&M model is that the new simulation uses the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986a). This algorithm makes use of pools of hidden units to capture nonlinearities in problem spaces. The improved learning of the training set found in the new simulation is at least partly attributable to this more powerful learning algorithm.

Training consisted of 24,000 runs or epochs, during which the network was trained to produce the correct output for various forms. Forms were presented with the actual relative frequencies found in (Francis, & Kucera, 1982), with the most frequent form being produced exactly once per epoch, and the rarest forms about once every 700 epochs. This resulted in the presentation of about 1.3 million forms during the course of training.

The input corpus was derived from the Francis and Kucera (1982) corpus of English word frequencies. The 6949 most frequent verb forms - including present, past, past participle, present participle and third person singular - were first selected as a base set for the simulation. These were derived from 2161 different verbs. Homophones and multiple forms (e.g. past tense for "spit" can be either "spit" or "spat") were eliminated by extracting the less common forms. All forms which had more than three syllables, more than three consonantal phonemes in a row, or more than two vocalic phonemes in a row were also removed. The remaining 6090 forms, derived from 2062 verbs, represented the corpus of forms used in the simulation. Of these the least frequent 10% of the regulars and the least frequent 10% of the irregulars were extracted and saved for testing the generalization abilities of the network. The remaining 5481 forms were used for training. The training set included 118 irregular past tense forms.

3. The Results

The most relevant results are those for the past tense. First, let us consider some general aspects of the irregular past tense results.

1. Typically, the present tense was produced as output during the first 40 to 80 epochs.
2. For zero-marking irregulars such as "cut" or "hit," the period of direct reproduction extended as far as epoch 200 before the first attempts were made to produce overregularizations such as "cutted" or "hitted."
3. For forms with vowel changes, the network spent a period from about epoch 200 to about epoch 3000 exploring alternative vowel shifts and modifications.
4. Concurrently with explorations of vowel shifts, the network explored use of final /t/, /d/, and /id/.
5. For most of the forms, correct performance was stabilizing by epoch 3000. However, a few late vowel errors and overuses of final /t/, /d/, or /Id/ can be found up to around epoch 6000.
6. In most cases, the last 10,000 epochs are error free.
7. There were three forms that were only learned in the last 10,000 epochs. These were "bled" at 14780, "brought" at 15810 and "thought" at 15820.
8. Only eleven low frequency irregular forms remained unlearned at the end of training.
9. In the testing done at the end of the simulation, nine of the thirteen *untrained* past tense irregulars were missed. Six of these were simply mistaken as other classes of irregulars or as regulars, as the network had no way to know for sure to which class they belonged.

By the first test point the regular past tense was already over 99% correct. In overall terms, the simulation obviously did very well at its task of learning the past tense. These results are encouraging, but how well was the model able to address the 13 specific problems we examined earlier?

3.1. The u-shaped learning problem

The u-shaped learning for verbs involves four observable components -- early correct usage, subsequent overregularization, coexistence of correct and incorrect forms, and final correct usage. The new simulation correctly models the last three components.

1. **Overregularization.** Usually, overregularization begins shortly after the first 50 or 100 epochs. For nearly every one of the irregular verbs, except for zero-marked pasts such as "hit" or "cut," there is a clear early period of overregularization. This is not an across-the-board phenomenon, since some verbs show much more overregularization than others and some show it for a much longer period than others.
2. **Coexistence of correct and incorrect forms.** There is then a period of between 300 and 3000 epochs for almost all irregular verbs where the correct form coexists with overregularized forms. For a verb like "bent" this period ends by epoch 2800; for a verb like "arise" it extends to epoch 8010.
3. **Final correct usage.** For nearly all of the verbs, there then follows a long period of final correct usage. It is important to note that, once the network reaches correct performance, deviations from the correct model are fairly infrequent. There are indeed late occurring errors against a background of correct performance, such as the uses of "binded" for "bound" on epochs 1950 and 2300. However, when we realize that the learning went on for 24000 epochs, it is clear that much of this period was spent in error-free performance for this verb. In this sense, the model behaves fairly deterministically, reflecting the deterministic nature of the learning algorithm. If we were interested in simulating adult speech errors, we would have had to employ a non-deterministic model.

However, the simulation fails to capture the fourth component of the u-shaped pattern -- early correct usage for irregular forms.

3.2. The "ated" problem

P&P argue that forms such as "ated" arise when the child confuses "ate" with the present. Capturing this type of processing in this network was extremely easy. We simply constructed a second set for testing generalization that was composed of four irregular past forms. The verbs were "ate," "broke," "ran," and "bought." The results are not particularly surprising. The network produced as output the forms "ated," "broked," "ranned," and "boughted." This was the case for "ate," "ran," and "brought" at the first test point (epoch 4000). "Broke," however, yielded "broke" as its past tense in this particular type of generalization testing until the test at 12000 epochs, after which point it produced "broked."

3.3. The "hitted" problem

Like the R&M model, the new simulation produced fewer overregularizations for zero-marking verb like "cut" and "hit" than for other irregulars. Indeed the zero-marking class was learned quite quickly with very few errors. For example, "burst" had only three "bursted" errors, and only seven errors in total. The verb "cast" had only one "casted" error after epoch 1950; "cost" had none after epoch 1340, and "cost" had none after epoch 210. Perhaps the most dramatic evidence for the power of what Menn and MacWhinney (1984) called the "repeated morpheme constraint" is in the verb "cut" which had only three errors in 24000 epochs -- two of which were uses of the overregularized "cutted" on trials 170 and 210. When we compare this virtually error-free learning of "cut" with the difficulties the network had in learning an alternation such as "deal - dealt," it is clear that verbs of the "hit" and "cut" type are much easier to learn than other irregulars.

3.4. The "algalgal" problem

The "algalgal" problem disappears in the new simulation. The phonological representation used in the new simulation guarantees unique representations for different sounding words. The left-justified representation of Oykangand "algal" is CCCaVlgCaVlCCVVCCC, whereas that of "algalgal" is CCCaVlgCaVlgCaVlCC. Since these different words have different representations, it is clear that the "algalgal" problem is solved by the new simulation.

3.5. The "slit-silt" problem

The "slit-silt" problem also disappears with the new simulation, because phonologically similar words now have more similar representations. In the case of "slit" and "silt" the /i/ appears in exactly the same slots for both words. The final /t/ also appears in the same slot in the right-justified representation.

3.6. The "brag-grab" problem

We ran an auxiliary simulation identical to the basic simulation, but in which the past tense was formed simply by reversing the order of segments in the present tense. For example, the past tense of "brag" was "grab" and the past tense of "trickle" was "lkirt." This transformation greatly impaired the performance of the network on the past tense. After 24000 epochs of training, only 15% of the forms were correct. Apparently, the network cannot learn an alternation of this type.

3.7. The phonological regularities problem

The new simulation addresses the phonological regularities problem in a variety of ways. First, the change to a feature/slot representation improves its ability to model generalization across the phonological inventory. To demonstrate this, the simulation was also given the verb "bach" /bax/ as a generalization test and correctly produced the past tense form "bached" /baxt/. Second, by including multiple forms of each verb in training, the simulation was able to demonstrate cross-paradigm regularities. For example, the progressive of "accompanying" was occasionally formed as "accompanyng." The deletion of the initial /I/ of the progressive -ing suffix appears to be a result of paradigmatic pressure from similar deletion patterns occurring in the other suffixes. Of course, the network learns to counteract this pressure, but the important fact is that the new simulation shows the presence of such phonological regularities. Other evidence of the impact of phonological regularities is provided by interpenetrations of past participle forms into irregular pasts.

3.8. The homophony problem

A small auxiliary simulation was run to demonstrate the ability of the network to acquire past tense forms for homophones. The verbs used were "ring," "wring," "jump," "want," and "run." The network architecture was supplemented by the following 21 semantic features: action, auditory-result, cause-contact, circle, completive, high-pitch, internal-state, object-gap, object-state, object-thing, positional-change, response, sharp-onset, speech-act, surround, torque, use-of-hands, use-of-feet, vertical direction, volitional, and whole-body-motion. The results of this simulation were quite simple. The network learned to produce "rang" as the past tense of "ring-1," "ringed" as the past tense of "ring-2," and "wrung" as the past tense of "wring" within 2400 epochs. These results indicate that nets of this type can readily resolve homophony.

3.9. The convergence problem

By epoch 4000, all but seven of the regular past tense forms were being produced correctly. By epoch 8000, all but one of the regular past tense forms were correct. By epoch 16,000, all of the regular past tense forms were correct. By the end of training at epoch 24,000, errors were being made on only 11 of the irregular pasts. Perfect performance on the present progressive was achieved by the first check at 4000 epochs. The third person singular present was perfected by the end of the simulation. Learning of the past participle followed a pattern similar to that for the past. It is clear that the network succeeded at its assigned task of learning the English verb paradigm. If we had allowed the network to run for several additional days or given it additional hidden unit resources, we probably could have reached complete convergence.

3.10. The regular pattern problem

Regular pasts were learned fairly early and without significant error. By the first check at epoch 4000, only seven of the 1059 reguar pasts were being missed. This dropped to one miss by the following check at epoch 8000, and by epoch 16000 performance was perfect. This type of learning is exactly what we see in children.

P&P suggest that, in some way, connectionist models are misportraying the role of the regular pattern. Before this issue can be seriously examined, P&P will have to show how this issue can have real consequences for computational simulations. For the moment all we can say is that the results of the new simulation for the regular pattern look just like what we see in real children.

3.11. The cryptorule problem

L&B criticized R&M for biases in the selection and construction of Wickelphones. In general, these complaints were reasonable. However, L&B attempted to move from these reasonable complaints to the general claim that acceptance of a particular featural representation in a PDP network is tantamount to acceptance of a production system architecture. There is no reason to accept this linkage, particularly in regard to the current simulation. More specifically, there is nothing in the featural representation of the current simulation that biases it toward the acquisition of some particular rule. Instead, the representation was chosen to be powerful enough to facilitate learning of different types of rules in different languages. It is true that the representation is capable of expressing something like "final t." However, there is nothing wrong with that ability and nothing that can be construed as involving either TRICS or "tricks."

3.12. The early noise problem

The new simulation no longer spends its initial epochs gropping toward the ability to produce some recognizable form of the verb. From the first trials, it is producing either the present tense or some variant of the present tense. In this regard the model provides a more accurate characterizaiton of the way in which the two-year-old child works on this task.

3.13. The direct access problem

The new simulation does not solve the direct access problem.

4. Is there a better symbolic model?

If there were some other approach that provided an even more accurate characterization of the learning process, we might still be forced to reject the connectionist approach, despite its successes. The proper way of debating conceptualizations is by contrasting competitive implementations. To do this in the present case, we would need a symbolist implementation that can be contrasted with the current connectionist implementation. P&P sketch out a piece of such an alternative but advise the reader not to take it as a "serious model." Unfortunately, the reader has no alternative but to take it seriously. It turns out that the model which P&P are proposing as the main symbolist alternative to the R&M model is fragment of the model of MacWhinney (1978). We are not arguing that the MacWhinney model was ill-constructed and descriptively inadequate. But when it comes to actually implementing the account of MacWhinney (1978), a myriad of detailed decisions must be made regarding the shape of possible input forms, the ways in which rule strengths should be incremented, the algorithm for production matching and conflict resolution, and so on. The current approach avoids these ad hoc modelling decisions.

5. Summary

We have examined a series of 13 problems raised by Pinker and Prince (1988) and Lachter and Bever (1988) against the verb learning model of Rumelhart and McClelland (1986). These problems were presented as fundamental flaws in the conceptualizations underlying connectionism that could well call into question any application of connectionist models to language processing or learning. To address these problems, we constructed a new simulation using a new input representation based on feature/slot units and two views on input words, a new architecture using identity mappings and hidden units, a new learning algorithm, and an input corpus that included all five cells of the English verb paradigm Together, these changes led to a vast improvement in the performance of the model. All of the 13 problems were addressed successfully except for those dealing with the problem of direct access within connectionist nets.

As a result of this work we conclude that connectionist models are indeed extremely useful ways of characterizing the learning inflectional systems. We also conclude that the critiques of Pinker and Prince and Lachter and Bever erred in confusing conceptualizations with implementations.

REFERENCES

Lachter, J., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. Cognition, 28, 195-247.

MacWhinney, B. (1978). The acquisition of morphophonology. Monographs of the Society for Research in Child Development, 43, Whole no. 1.

Menn, L., & MacWhinney, B. (1984). The repeated morph constraint: Toward an explanation. Language, 19, 519-541.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing Model of language acquisition. Cognition, 29, 73-193.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by back propagation. In D. Rumelhart, & J. McClelland (Ed.), Parallel Distributed Processing: Explorations in the microstructure of cognition. Cambridge, MA: M.I.T. Press.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, & D. E. Rumelhart (Ed.), Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MIT Press.