Are Rules a Thing of the Past? The Acquisition of Verbal Morphology by an Attractor Network.

James Hoeffner
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
jh6s@andrew.cmu.edu

Abstract

This paper investigates the ability of a connectionist attractor network to learn a system analogous to part of the system of English verbal morphology. The model learned to produce phonological representations of stems and inflected forms in response to semantic inputs. The model was able to resolve several outstanding problems. It displayed all three stages of the characteristic U-shaped pattern of acquisition of the English past tense (early correct performance, a period of overgeneralizations and other errors, and eventual mastery). The network is also able to simulate direct access (the ability to create an inflected form directly from a semantic representation without having to first access an intermediate base form). The model was easily able to resolve homophonic verbs (such as *ring* and *wring*). In addition, the network was able to apply the past tense, third person *-s* and progressive *-ing* suffixes productively to novel forms and to display sensitivity to the subregularities that mark families of irregular past tense forms. The network also simulates the frequency by regularity interaction that has been found in reaction time studies of human subjects and provides a possible explanation for some hypothesized universal constraints upon morphological operations.

Introduction

In recent years the status of rules in cognitive science has become an issue of heated debate. Many cognitive scientists believe that explicit rules are necessary to explain human behavior (Pinker & Prince, 1988; Lachter & Bever, 1988). Others have challenged this reliance on rules and symbolic systems (Rumelhart & McClelland, 1986; MacWhinney & Leinbach, 1991).

At the center of this debate has been the study of inflectional morphology; the acquisition of the past tense has been of particular importance. Researchers have found that the acquisition of the English past tense involves three distinct stages (Kuczaj, 1977, 1978; MacWhinney, 1978). In the first stage, children correctly produce a small number of both regular and irregular forms. In the second stage they sometimes "overregularize" irregular forms, producing errors such as *goed* and *ated*. Particularly striking is the fact that children will sometimes overregularize irregular forms that they had previously produced correctly. In the third stage, which is only reached gradually over a period of years, the children exhibit total mastery of both regular and irregular forms with very few errors of any kind.

The existence of both overregularizations and a U-shaped curve have been considered strong evidence for the belief that language learning involves organizing linguistic knowledge into a system of rules and exceptions to those rules. Rumelhart & McClelland (1986) challenged this belief by creating a connectionist network which, they claimed, simulated the process of learning the English past tense without making use of any explicit rules. Several authors (Pinker & Prince, 1988; Lachter & Bever, 1988) criticized the Rumelhart and McClelland account on a variety of grounds. These criticisms can be grouped into several clusters: 1) Problems with the phonological representation, 2) Criticism of the training regimen, 3) Problems stemming from the lack of a semantic representation in the model, 4) The failure to incorporate constraints on possible forms, 5) The failure to simulate the differential effects of frequency on regular and irregular forms. The criticisms raised in 1 and 2 have been addressed with some success by later connectionist models (MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991). This paper addresses the criticisms in 3, 4 and 5.

Problems Stemming from the Lack of a Semantic Representation

All previous connectionist models of past tense acquisition, with the exception of Cottrell & Plunkett (1991), have been phonology to phonology models. They took the verb stem as their input representation and converted it to the past tense (or in the MacWhinney & Leinbach model, to a variety of inflected forms). Because of this, these models were unable to display some elementary properties of natural languages. One of these properties is homophony. Homophones are pairs of words that sound identical but have different meanings. English and other languages have many homophones, for example, there are the verbs *ring* (a bell), *wring* (your hands) and *ring* (form a circle around). The phonology to phonology networks cannot learn to inflect verbs that have homophonic stems but different past tenses (*ring-rang* vs. *wring-wrung* or *ring-ringed*).

Another related problem is that of direct access. The phonology to phonology models take a base form and transform it into an inflected form. People seem to display an ability similar to this. But as MacWhinney and Leinbach point out, "...intuition, theory (MacWhinney, 1978) and experimentation (Stemberger & MacWhinney, 1978) all suggest that we can also access derived forms directly. In other words, we learn that *ran* means *running in the past* and use this knowledge to access *ran* directly without starting off at *run*." Phonology to phonology models cannot simulate this process of direct access.

There are at least two ways to address these problems within a connectionist framework. One way is to add a semantic representation to the input of a phonology to phonology network as MacWhinney and Leinbach did in a small subsidiary model. In principle, this approach can resolve the homophony problems but it does not address the direct access problem. Also, in practice, the net result of adding semantic representations to the input of a phonology to phonology network is to reduce the overall performance of the model, since the network must devote considerable resources to learning the largely arbitrary associations between semantic and phonological representations.

The approach taken in this paper is a different one. The network is trained to form associations between semantic and phonological representations. In this type of model, direct access is the basic process and the ability to inflect a novel form is a secondary ability parasitic on the direct access process. By taking this route, we may be able to solve both the homophony and direct access problems simultaneously and in a principled way.

Another criticism that can be addressed by using a semantics to phonology network is the question of "double-marked" forms like *ated* or *wented*. In the Rumelhart & McClelland model, these errors are produced by blending the specific *go->went* mapping with the regular stem->stem+ed mapping. But Pinker & Prince claimed that the fact that children occasionally produce errors like *wenting* and *ating* even when the progressive is fully regular shows that these errors are caused by feeding the irregular past into the regular suffixation process.

In MacWhinney & Leinbach's model, they were able to produce double-marked errors simply by feeding the past tense (e.g., *ate*) into the network as if it were the stem. In the present model we take a different approach, one based on the nature of the semantic relationships among verbs (Bybee, 1985). In our model, all the members of a particular verbal paradigm share an identical "core" semantic representation. The semantic representations of two members of the same paradigm (e.g., *jump* and *jumped*) are distinguished only by the relatively small number of units which code for inflections. Because of the proximity of the semantic representations of the members of a verbal paradigm it is possible for the network to produce phonological blends of various kinds including the combination of an irregular past form like *went* with a regular suffix like the progressive *-ing*.

Constraints on Possible Forms

Pinker & Prince raised an intriguing criticism of the Rumelhart & McClelland model. There seem to be logically possible inflectional devices that are never used in any language. For example, no language inflects a verb for tense by transposing all the phonemes in the word (making *tih* the past tense of *hit* or *pals* the past tense of *slap*). Another ubiquitous feature of natural language is preservation of the stem in inflected forms. This is true not only of regular forms but of most irregulars as well. In most irregulars the majority of the phonetic material in the stem is preserved in the inflected form (such as *sing-sang* , *give-gave*, etc.). Pinker & Prince claimed that Rumelhart & McClelland's model was insensitive to such universal constraints on possible morphological operations and therefore was invalid as a model of how people actually learn and represent language.

We believe that there is a strong bias against certain types of morphological operations and biases in favor of other types of operations (for example, affixation of a stem or base form). At least some of these biases may result from the nature of semantic level relationships among different verb forms. For example, in our model, the phenomena of preservation of the stem in inflected forms can be explained in the following way. The semantic representations of the members of an inflectional paradigm are very close. All things being equal, connectionist models prefer mappings where the similarity structure of the output set reflects the similarity structure of the input set. This built in bias means that the network will find it much more difficult to output *hit* as the present tense and *tih* as the past then it would be to learn a verbal paradigm where much of the stem was preserved in every inflected form.

Differential Frequency Effects

Prasada, Pinker & Snyder (1990) tested subjects' reaction times in a past tense generation experiment. They found that higher frequency irregulars were produced faster than low frequency irregulars but the same frequency effect was not found for regular past tense forms. They claim that this evidence supports the qualitative distinction between regular and irregular inflected forms. The regulars are stored in an associative memory device that is sensitive to frequency and similarity but the regularly inflected forms are not. Only the stems need to be stored, since there is a second mechanism, an affixation procedure that can take any stem and append the *-ed* suffix to it.

Seidenberg and Bruck (1990) found a similar frequency by regularity interaction in the production of past tense forms, but gave the results a different interpretation, one that is consistent with the behavior of a single connectionist model that learns to produce both regular and irregular verbs. With the attractor network presented in this paper, we can simulate some of the effects of frequency and regularity on reaction time and show that the interaction need not be taken as evidence for a dual mechanism account.

The Model

The network presented in this paper is an attractor network (Hinton & Sejnowski, 1986; Hopfield, 1984). Attractor networks have recurrent connections; these enable the network to develop stable resonant or attractor states. There are at least three advantages to using an attractor network.

1. Learning arbitrary associations: The mapping from semantics to phonology is difficult because the two representations are arbitrarily related. Feed-forward networks need large weights and prolonged training to learn such a mapping. But, as Plaut & Shallice (1991) point out: "They [attractor networks] are also more effective at learning arbitrary associations because the reapplication of unit non-linearities can magnify initially small state differences into quite large ones. "
2. Flexibility: We can interrogate the knowledge stored in the network by presenting it with pieces of information and allowing the network to fill in the rest of the learned pattern. This characteristic of attractor networks will enable us to probe the network's ability to use its knowledge of morphology productively in ways that could not be done with feed-forward or simple recurrent networks.
3. Reaction time measure: We can use the network's settling time as an analog of reaction time in psychological experimentation and simulate the effects of frequency and regularity on subjects' reaction times.

Network Architecture: The network contains 185 units arranged in three layers: the semantic, hidden and phonological layers. The semantic layer contains 96 units, the hidden layer has 68 units and the phonological layer has 21. The semantic layer has bi-directional connections to the hidden layer and the hidden layer has bi-directional connections to the phonological layer. All three layers are fully intra-connected (each unit in a layer is connected to every other unit in that particular layer).

Semantic representation: In the semantic layer, 84 of the 96 units are used to represent the "core" meaning of each verbal paradigm. 12 more units are dedicated to representing inflectional markings. The semantic representation of a verb consists of a randomly generated pattern over the 84 "core" units plus the 12 inflectional units. The semantic representations of the members of an inflectional paradigm (e.g., *jump, jumped, jumps and jumping*) will have an identical "core" representation. The only difference will be in the 12 units that code the inflections. Different verbs (e.g., *jump* and *walk*) will have "core" semantic representations that must differ by at least 24 of the 84 units. This ensures that the semantic distance between different verbs will be greater than the semantic distance between two members of the same paradigm.

Phonological representation: The phonological representation was based on an artificial language devised by Plunkett & Marchman (1991). In this language each phoneme is uniquely represented by a 6 bit vector. Each word is made up of three phonemes. The words can have a CVC(hit), CCV(try) or VCC(ask) syllabic structure. There are also three units in the phonological layer that code for the inflectional suffixes.

Training corpus and schedule: The corpus consists of 800 verb forms (200 paradigms). Each paradigm consists of an unmarked "stem" and three inflected forms: the past tense, the third person singular, and the progressive. The 3s and progressive forms are totally regular, the only irregulars are past tense forms. There are four types of past tense forms in the artificial language. Regular verbs, Vowel Change(VC) verbs, No Change(NC) verbs, and Arbitrary verbs. The past of the Regular verbs is the stem + ed. The past tenses of the Vowel Change verbs are formed by altering features of the vowel. The past tense of the No Change verbs is identical to the stem. And the past tense of the arbitrary verbs bear no systematic phonological relationship to the stem. These four types correspond to the predominant types of stem to past mappings found in natural languages. For example, in English, *go-went* is an Arbitrary, *put-put* is a No Change, and *give-gave* is a Vowel Change. There are 20 irregular pasts in the training set, 2 arbitraries, 6 no change and 12 vowel change. The vowel change verbs are in two clusters, in VC_1 all the stems have the form the C+ing and their pasts have the form C+ang. The VC_2 verbs are conjugated in the following way: C+aif ->C+If.

Both the regular and irregular verbs are organized into high, medium and low frequency groups. For the irregulars there are 8 high, 6 medium and 6 low frequency verbs. For the regulars, there are 40 high, 100 medium and 40 low frequency verbs.

The most frequent inflected forms are presented 9 times per epoch. The least frequent are presented only once. For a given paradigm, all three inflected forms have the same frequency. In all paradigms, the stem is presented twice as often as any of the inflected forms. For the entire corpus, the regular to irregular *type* frequency ratio is 9:1, but the regular to irregular *token* frequency ratio is less than 3:1 because of the higher frequencies of the irregular verbs.

Training schedule: The network is first presented with 30 paradigms, 20 regular and 10 irregular. The network is trained for two epochs on this subset. Then the corpus is expanded incrementally with 4 new paradigms added each epoch up till epoch 45. At no point is there a sudden influx of regular verbs or a rapid change in the relative proportions of regular and irregular verbs.

Learning Algorithm: The network was trained with the Contrastive Hebbian algorithm (Peterson & Anderson, 1987).

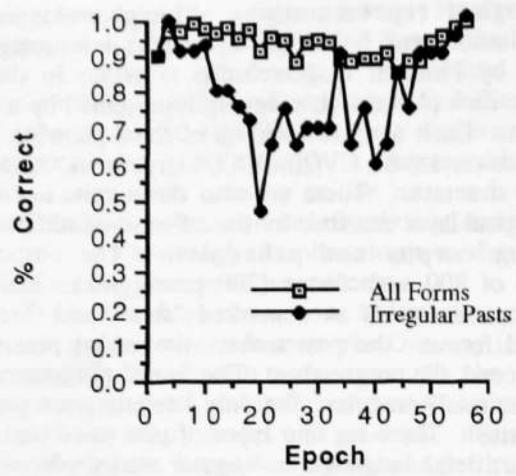

Figure 1: Percent correct on all forms and irregular past tense forms only.

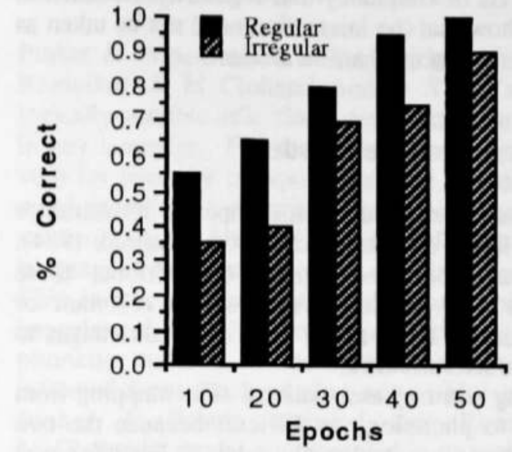

Figure 2: Past tense inflection of novel forms with regular and irregular phonology.

Results

Overall Performance: The network had no difficulty learning the task. Initially, performance was very strong, reaching the high 90's by the 4th epoch. And even though the network was absorbing 16 new forms per epoch, its performance never dipped below 85% correct. Across all 50 epochs, the network averaged 93% correct. At the end of training (epoch 58) the network was producing the 800 verb forms with 98% accuracy. (See figure 1 and Table 1).

U-shaped curve: The network displayed all three stages of the U-shaped developmental pattern. From the beginning of training till epoch 10, the network averaged 93% correct on irregular pasts. During this period, it did not make any overregularization errors. At epoch 12, the network began to overregularize the past tense suffix and its performance on irregular pasts fell to 73%. The network continued to overregularize the irregular pasts throughout most of the training period. By epoch 44, the number of overregularization errors began to diminish as the network learned how to inhibit the suffix when producing an irregular past. By the end of training, the network achieved perfect performance on the irregular pasts, thereby completing the third phase of the U-shaped curve.

In Table 2, all the overregularization errors made by the network are listed. The network made all three possible types of overregularization errors, Base+ed (goed), Past+ed (wented) and No Change+ed (hitted). The network displayed several of the characteristics of a *micro*-U-shaped developmental curve (Plunkett & Marchman, 1991). There was never an across the board overregularization of all or even most of the irregulars. Instead, some verbs were overregularized frequently and others not at all. In addition, overregularization was highly variable for many verbs. For example, No Change Verb #3 was produced correctly till epoch 14,overregularized on epoch 16, was correct on 18 and 20, incorrect again on epoch 22, correct again on 24 and 26 and overregularized once more on epoch 28.

Homophones: There were two pairs of homophones in the training set. Both homophonic pairs were mastered by the network.

Double-Marked Errors: Most of the network's overregularizations were Base+ed(goed) or No Change+ed(hitted) but the network did produce several "double-marked" Past+ed(wented) errors. It also produced a small number of errors similar to "wenting", by combining the irregular past with the 3s or -ing suffix.

Productivity and sensitivity to subregularities: Figure 2 shows the model's ability to generalize its knowledge and apply the three inflectional suffixes correctly to novel forms. The network was presented with a corpus of 40 novel stems. Half were phonologically similar to the irregular verbs the network had been trained on. The other 20 novel verbs did not resemble the irregulars. The network was tested in the following way: the semantic layer inflectional units are "soft-clamped". The novel stem is also "soft-clamped" to the phonological layer. The 3 suffix units in the phonological layer and the rest of the network are left free. The task of the network as it settles is to properly activate the suffix units, thereby inflecting the novel stem. This task is analogous to probing a human subject's ability by presenting him with a novel stem like "glorp" and asking what the past tense of it would be.

The network was able apply all three suffixes productively. Figure 2 shows that the network was able to reach perfect performance in applying the past tense suffix to the phonologically regular stems by epoch 50. This compares very favorably with the generalization performance of Cottrell & Plunkett's(1991) semantics to phonology network, which was was unable to achieve greater than 55% correct generalization to novel stems.

The network also showed sensitivity to the subregularities that mark families of irregulars. The irregular phonology novel pasts were consistently less likely to be suffixed than the regular phonology novel pasts. This shows that the network has learned the phonological regularities that mark irregular families (such as the fact that all the No Change verbs end in t/d) and can use this information to block the regular suffixation process.

The network was also able to apply the -s and -ing suffixes productively as well (see Figure 3). In fact, because they are totally regular, the network learned to use these suffixes more quickly than the it did for the past tense (80-85% correct after only 20 epochs). Also, for these suffixes, there was no consistent effect of irregular phonology. The network was able to inflect novel irregulars as easily as novel regulars.

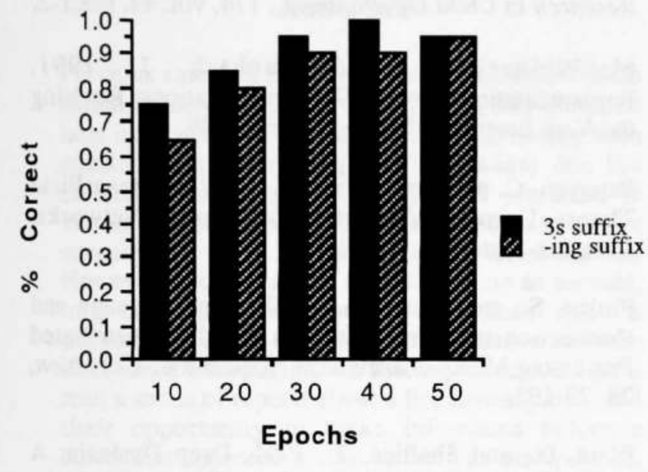

Figure 3: Inflection of novel forms with -s and -ing suffixes.

Frequency Effects: We can use the network's settling time as an analog of reaction time in psychological experimentation. In reaction time experiments the subject is presented with a base form (walk) and told to generate the past tense (walked) as quickly as possible. We can simulate this effect by "soft-clamping" a stem to the phonology layer at the same time that we clamp the full semantic pattern for the verb's past tense. The network has to convert the soft clamped stem to the proper past tense form.

Figure 4 shows that the network also displays the frequency by regularity interaction that has been found in studies of human subjects. (Only correct responses were used, this removed ~2% of the responses). For the irregulars there is a pronounced frequency effect, but for the regular verbs, there is very little effect of frequency.

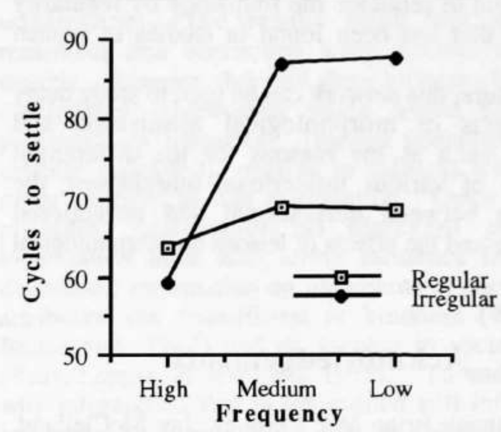

Figure 4: Frequency by regularity interaction

Epoch	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
All forms	90	97	97	99	98	96	98	95	97	91	95	94	88	93	94
Stems	100	100	100	100	98	100	100	96	99	96	98	94	90	98	96
Past tense	63	89	89	96	97	92	93	90	93	85	88	90	85	90	91
3s	97	100	100	100	100	97	100	98	99	90	98	96	88	93	94
-ing	97	100	95	100	98	97	100	95	98	94	96	95	89	92	93
Irregular past	90	100	92	92	93	80	80	76	72	47	65	75	65	70	70

Epoch	32	34	36	38	40	42	44	46	48	50	52	54	56	58
All forms	93	89	88	89	89	90	85	88	88	93	94	94	96	98
Stems	97	92	92	91	89	89	90	90	92	96	96	96	97	97
Past tense	86	87	85	82	81	84	85	88	87	89	93	94	94	98
3s	97	92	89	88	92	92	80	88	88	93	94	95	97	98
-ing	93	86	87	94	89	92	85	87	86	95	94	95	97	98
Irregular past	70	90	65	75	60	65	85	75	90	90	90	95	90	100

Table 1: Percent correct by epoch.

Epoch	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	Totals
Base+ed	1	2	2	1		3	1	3	3	2	1		1	1	3	3	1			28
NoChange+ed	1	1	1	2	2	3	2	2	2	1	2		3	1		2	1	2	1	29
Past+ed					2	1	2						1							6
Totals	2	3	3	3	4	7	5	5	5	3	3	0	5	2	3	5	2	2	1	63

Table 2: Distribution and types of overregularization errors.

Discussion

Our network was able to show the most salient aspects of English past tense acquisition. The network passed through all three stages of the U-shaped developmental curve and did so without the need for a sharp input discontinuity. The model demonstrated the two behaviors that were thought to be paradigmatic of rule acquisition and use (overgeneralizations and *micro* U-shaped development). By using a semantics to phonology attractor network, we were able to solve the direct access, double-marking and homophony problems. We were also able to shed light on the basis for certain universal morphological biases such as preservation of the stem in inflected forms and the constraint against mirror image morphological mappings. Furthermore, we were able to replicate the frequency by regularity interaction that has been found in studies of human subjects.

In the future, this network can be used to study many other aspects of morphological acquisition and processing such as the reasons for the differential acquisition of various inflectional morphemes, the differences between derivational and inflectional morphology and the effects of lesions on morphological processing.

Acknowledgements

I'd like to thank Brian MacWhinney, Jay McClelland, Randy O'Reilly and Priti Shah for their aid and advice.

References

Bybee, J.L., 1985. *Morphology: A study of the relation between meaning and form.* Philadelphia: Benjamins.

Cottrell, G. W., and Plunkett, K. , 1991. Learning the Past Tense in a Recurrent Network: Acquiring the Mapping from Meaning to Sounds. In Proceedings of the Thirteenth Annual Meeting of the Cognitive Science Society.

Hinton, G.E., and Sejnowski, T.J., 1986. Learning and Relearning in Boltzmann machines. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group (Eds.)., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* (Vol. 2). Cambridge, MA: Bradford Books.

Hopfield, J. J., 1984. Neurons with Graded Responses have Collective Computational Properties like those of Two State Neurons. In Proceedings of the National Academy of Science, U.S.A., 81, 3088-3092.

Kuczaj, S. A., 1977. The Acquisition of Regular and Irregular Past Tense Forms. *Journal of Verbal Learning and Verbal Behavior*, 16, 589-600.

Kuczaj, S. A., 1978. Children's Judgments of Grammatical and Ungrammatical Past Tense Forms. *Child Development*, 49, 319-326.

Lachter, J., and Bever, T.G., 1987. 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* , 174, vol. 43, nos.1-2.

MacWhinney, B. and Leinbach, J., 1991. Implementations are not Conceptualizations: Revising the Verb Learning Model. *Cognition*, 39.

Peterson, C. and Anderson, J.R. ,1987. A Mean Field Theory Learning Algorithm for Neural Networks. *Complex Systems* 1, 995-1019.

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

Plaut, D., and Shallice, T., 1991. Deep Dyslexia: A Case Study of Connectionist Neuropsychology. Unpublished manuscript.

Plunkett, K., and Marchman, V., 1991. U-shaped Learning and Frequency Effects in a Multilayered Perceptron: Implications for Child Language Acquisition. *Cognition*, 38, 43-102.

Prasada, S., Pinker, S., and Snyder, W., 1990. Some Evidence that Irregular Forms are Retrieved from Memory but Regular Forms are Rule Generated. Paper presented at the 31st annual meeting of the Psychonomic society (November: New Orleans).

Rumelhart, D. E., and McClelland, J.L., 1986. On Learning the Past Tense of English Verbs. In D.E. Rumelhart, J.L. McClelland and the PDP Research Group (Eds.)., *Parallel Distributed Processing: Explorations in the microstructure of cognition.* (Vol. 2). Cambridge, MA: Bradford Books.

Seidenberg, M.S., and Bruck, M., 1990. Consistency Effects in the Generation of Past Tense Morphology. Paper presented at the 31st meeting of the Psychonomic Society (November: New Orleans).

Stemberger, J., and MacWhinney, B., 1985. Frequency and the Lexical Storage of Regularly Inflected Forms. *Memory and Cognition*, 14, 17-26.