COGNITIVE SCIENCE Vol 23 (4) 1999, pp. 463–490 ISSN 0364-0213
Copyright © 1999 Cognitive Science Society, Inc. All rights of reproduction in any form reserved.

A Connectionist Model of English Past Tense and Plural Morphology

KIM PLUNKETT
PATRICK JUOLA
University of Oxford

The acquisition of English noun and verb morphology is modeled using a single-system connectionist network. The network is trained to produce the plurals and past tense forms of a large corpus of monosyllabic English nouns and verbs. The developmental trajectory of network performance is analyzed in detail and is shown to mimic a number of important features of the acquisition of English noun and verb morphology in young children. These include an initial error-free period of performance on both nouns and verbs followed by a period of intermittent over-regularization of irregular nouns and verbs. Errors in the model show evidence of phonological conditioning and frequency effects. Furthermore, the network demonstrates a strong tendency to regularize denominal verbs and deverbal nouns and masters the principles of voicing assimilation. Despite their incorporation into a single-system network, nouns and verbs exhibit some important differences in their profiles of acquisition. Most importantly, noun inflections are acquired earlier than verb inflections. The simulations generate several empirical predictions that can be used to evaluate further the suitability of this type of cognitive architecture in the domain of inflectional morphology.

I. INTRODUCTION

The acquisition of past tense in English has been long studied as a general touchstone for the development of morphology and productive linguistic rules in children. The general pattern of past tense formation for English is well understood; the overwhelming majority of English verbs have a simple past tense form which can be described as the addition of one of three variants of the "-ed" suffix to a base stem. A significant minority, particularly of relatively common verbs, take a so-called "irregular" form, which may or may not be systematically related to the stem form or to the forms of other words. The developmental course is also well understood; children typically begin by correctly producing a small

Direct all correspondence to: Kim Plunkett, Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK; E-mail: ⟨plunkett,patrick⟩@psy.ox.ac.uk.

number of both regular and irregular forms, then produce characteristically “over-regularized” forms for a small but significant fraction of their verb forms. They then appear to re-learn the correct form, producing the classic “U-shaped developmental profile” (Berko, 1958; Ervin, 1964; Kuczaj, 1977; Marcus et al., 1992).

Interpretations and models of this phenomenon vary; Marcus et al. (1992) have suggested (see also Pinker & Prince, 1988; 1991; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995) that the emergence of the over-regularized forms is indicative of the development of “a mental operation implementing the *-ed*-suffixation rule posited by grammarians” [pg. 8]. According to this theory, there are two routes to produce a past tense; either by reproduction of a memorized (irregular) form, or by applying a general rule to any word-form not recognized as being one of the forms in memory. Other researchers (Daugherty & Seidenberg, 1992; Rumelhart & McClelland, 1986; Plunkett & Marchman, 1991; MacWhinney & Leinbach, 1991) have argued instead that a single connectionist network is capable of producing appropriate patterns of behavior, and thus that a single associative route suffices to explain the evidence (see Christiansen & Chater, this issue, for an overview).

English noun plurals share many of the same characteristics as verb past tenses. There is a similar general rule describing most forms and a small, semi-regular group of common exceptions. Brown (1973), Marcus (1995), Marchman, Plunkett, and Goodman (1997) have described broadly similar time courses for the acquisition of plural nouns, including the U-shaped curve and approximately similar overall rates of over-regularization. Likewise, many of the same phonotactic features (such as voicing assimilation and/or epenthesis) are relevant to the acquisition of both noun and verb morphology. There are some crucial differences, however. The number of irregular noun types is nearly an order of magnitude smaller (about 20 irregular nouns versus about 150 irregular verbs), but they are individually more frequent. Nouns themselves are more frequent in running text, when measured in terms of either token or type frequency. Nouns are also less complex in the (irregular) inflections that they undergo. For example, all noun plurals share their onsets with the singular form (unlike the present/past pair “go”/“went”), and many irregular nouns involve a simple change in voicing of the final consonant from /f/ to /v/ (as in “thief”/“thieves”). By contrast, verbs participate in a wider variety of irregular paradigms including some verbs that appear to bear no relationship to their inflected forms. There also are some subtle developmental psycholinguistic differences. For example, noun plurals are typically learned more quickly than verb past tenses (Brown, 1973; Marchman et al., 1997). Similarly, over-regularization of noun plurals is likely to be observed earlier and more frequently in development than over-regularization of past tense forms (Marcus, 1995; Marchman et al., 1997).

These similarities and differences suggest that modeling the interactions between the acquisition of noun and verb morphology may also reveal important insights into the nature of the mechanisms underlying the cognitive architecture of these two inflectional processes. For instance, is it necessary to posit separate inflectional routes for regular nouns and regular verbs? Do irregular nouns and verbs inhabit the same associative memory system? How can the functional separation between irregular and regular forms be best explained?

A related question concerns the nature of the processing involved when the verb form itself is derived from a noun stem (as in denominal verbs). Kim, Marcus, Prince, and Prasada (1991) argue that denominal verbs are treated differently from identically sounding verb tokens in the inflectional process, suggesting that word meaning can be an important factor in inflection. On the other hand, evidence from slips of the tongue (Fromkin, 1973; Garrett, 1980) suggests that the process of inflection can occur independently of the intended sentence meaning (e.g., Fromkin's "days /deIj+z/ of the week" becomes "weeks /wijk+s/ of the day," accommodating to the new location for the plural inflection). Garrett has even demonstrated a case where the irregularization process is automatic and the (incorrect) word token takes an irregular inflected form in place of a regular token ("I'd know one if I heard it" becoming "I'd hear one if I knew it"), suggesting that accommodation and irregularization are automatic and independent of meaning. This argues for a functional separation between the syntactic process of inflection and the semantic process of the meaning incorporated by the inflectional process which occasionally separates as in the above errors.

One obvious weakness of the "dual-route" model is its inability to generalize to multiple paradigms cleanly. A word may be irregular (and thus a memorized exception) with respect to one syntactic form but not others—"go" takes an irregular past tense, but its plural is the regular "goes" instead of *"wents". Some nouns derive their inflected denominal forms from the singular (e.g., "knife" "to knife" "knifed"), while others from the plural (e.g., "half" "to halve" "halved"). These variations cannot be explained without resorting to further rules and a detailed (and complex) theory of the timing of rule application. Similarly, dual-route models assume system-general rules to account for the processes of voicing assimilation and epenthesis in inflected forms. However, dual-route accounts offer no motivation as to why these rules should be identical for nouns and verbs. The single-route model has the potential to provide a natural explanation in that noun and verb inflections are accommodated within the same system.

Here we present a single-system, feed-forward, connectionist model to compare the acquisition of noun and verb inflection head-to-head. To this end, we have constructed a model that simultaneously acquires noun plurals and verb past tenses. We use this model to determine the patterns of mastery (and errors) produced by a network which only has access to phonological representations of stems and their inflected forms, and their type and token frequencies. We demonstrate that the patterns of performance in the network are comparable with the acquisition data for young children.

In particular, the network exhibits a general advantage in acquiring noun morphology before verb morphology. Nevertheless, both regular and irregular nouns and verbs exhibit an early stage of error-free performance before over-regularization errors occur. Although the onset of over-regularization errors is earlier for nouns than verbs, errors in both categories are shown to be driven by critical mass effects (Marchman & Bates, 1994; Plunkett & Marchman, 1993). The network model also predicts a developmental shift in the relative ease of learning irregular nouns and verbs. Early in development, irregular nouns show an advantage but this advantage shifts as training proceeds.

We also examine the network model's behavior on denominal verbs and deverbal nouns. Both categories show a strong tendency to be regularized by the network. In order to deal with denominal verbs that are homophonic with irregular verbs, we introduce a supplementary set of simulations in which the phonology-to-phonology network is augmented with semantic inputs, thereby offering the network the opportunity to inflect phonologically identical forms in different ways. Previous work has demonstrated that neural networks can be trained to produce homophonic phonological outputs given distinct semantic inputs (MacWhinney & Leinbach, 1991; Cottrell & Plunkett, 1994; Hare & Elman, 1995). Again, under these conditions, the augmented network shows a strong tendency to regularize denominal verbs. These findings capture patterns of behavior reported elsewhere in the literature (Kim et al., 1991). Finally, we discuss the implications of these simulations for characterizing the cognitive machinery underlying past tense and plural morphology.

II. METHOD

Modeling Assumptions

We model the acquisition of inflectional morphology as the process of converting a representation of the stem of a word (in this work, more specifically, a phonological representation of the *sound* of the uninflected form of a word) to a representation of the sound of an appropriately inflected form, such as a past tense or a plural. Our focus is on the time course of development for acquisition, using the standard metaphor that increasing training epochs corresponds roughly to the increasing age and experience of the child. The system is presented with input/output pairs (stems and their inflected forms) and is allowed to operate on its internal weight states to reduce the difference between the actual output (of the current weights) and the desired output.

In this system, the child's processing is viewed as a *hypothesis generator* which runs concurrently with the normal tasks of production and comprehension, and can be functionally described as the child's attempt to systematize the pre-existing representations of inflectional sets (such as stem/singular/plural or stem/present tense/past tense). We assume that the child already has some form of a mental lexicon, including (at a minimum) a representation of the phonology of a word, plus some concept of syntax and semantics describing the word's pattern of use. The child is continually taking in word tokens and comparing the words actually heard (e.g., “went”) to the tokens that the child's hypothesis generator would have expected to produce as inflected forms of a given stem; when they differ, this provides evidence to the child that the hypotheses are wrong and should be modified. The hypothesis generator comes into its own when the child produces words, as the generator describes the current best guess (to the child) of how words ought to be inflected, especially in the case of nonce words for which the child has no semantics, little syntax, and no lexical entry.

Our model therefore hinges on some sort of a priori notion of “stem” available to the child (which is stable across different words) and a mental lexicon, which we assume also includes aspects of syntax and semantics that are not part of this modeling work. We assume further that the child is comfortable with different inflectional paradigms, and has the ability to analyze the sounds/phonemes of the language of interest correctly. We do not model the very earliest stages of language acquisition before these abilities are stably available to a child. Similarly, this work does not deal with the difficulties of correct analysis (determining whether or not a word is being used as a noun or a verb) or the difficulties of semantic acquisition.

Network Configuration

The connectionist simulations use a multi-layer perceptron network employing back-propagation of error (Rumelhart, Hinton, & Williams, 1986). For every input pattern presented to the system during training, the system calculates its current best guess of the relevant output pattern, which is then compared to the actual desired output pattern. Error is allocated to the output and hidden units proportionally with the RMS error and then weight changes are propagated backwards through the connecting layers. The simulation was built using the PlaNet simulator (Miyata, 1991) using 130 units for the input layer, 160 units for the output layer, and 200 units as the hidden layer. The system was trained with a learning rate (η) of 0.1 and no momentum ($\alpha = 0$). Weights were initially randomized in the range ± 0.5. Training was performed via a pattern update schedule, where each pattern was presented individually to the network (in random order) and training performed on each pattern.

Training Corpus and Representation

The training data for the simulations were taken from the CELEX corpus (Baayan, Piepenbrock, & Rijn, 1993); we extracted from this database all words which were monosyllabic, that contained no “foreign” sounds in their pronunciation according to the Moby Pronunciator database (Ward, 1997), and for which we had evidence that they could be used as nouns or verbs. This yielded a total corpus of 2626 stems, which encompassed 3226 total inflected word types (2280 nouns and 946 verbs). Of these word types, 26 were irregular nouns and 122 were irregular verbs. For these words, we took the corresponding token frequencies (of the stems) from the Brown corpus (Kučera & Francis, 1967) as a rough measure of token frequencies in running speech. These token frequencies were also heavily dominated by nouns; for the 17129 tokens in the training set, 13045 were noun tokens (204 of them irregular) and 4084 were verb tokens (997 of them irregular).

These numbers, although accurate, may justly be regarded with a certain degree of suspicion with regard to their appropriateness as a measure of a child’s input, as the Brown corpus (from which the frequency data derives) is a sample not of children’s (or child-directed) spontaneous speech, but of written, edited, adult-to-adult communication such as novels, magazines, and newspapers. On the other hand, measuring only child-directed or child-initiated speech could also be misleading as most children certainly listen to adult conversation (and even edited adult speech, e.g., on television). In the absence of accurate frequency counts of individual nouns and verbs for children, we have opted to use published frequency data. A further refinement of this work would employ frequency counts taken from computerised databases such as CHILDES (MacWhinney, 1991).

The training corpus was prepared by converting the Moby symbolic pronunciation (Ward, 1997) into a large binary vector using a modification of the PGPfone alphabet representation (Juola & Zimmermann, 1996), summarised in Table 1. Each phoneme was represented as a cluster of 16 binary phonetic features including aspects such as place, manner, and height of articulation, as described in Figure 1. Each word was divided into onset-nucleus-coda constituents and right-justified within a CCCVVCCC template (e.g., the word "cat" (/kAt/) would be represented by the training pattern ##k#A##t, the word "ox" (/Aks/) by ####A#ks, where "#" represents an absent sound). To this 128-bit pattern, two additional bits were appended representing the syntactic form to be inflected into, either the past tense (of a verb) or the plural (of a noun). The desired outputs were a similar encoding of the phonology of the inflected form, including an optional epenthetic vowel and final consonant. For instance, "cat" becomes "cats" (/kAts/), represented by ##k#A##t#s.

The token frequencies of words were individually tabulated as nouns and verbs, then the function $\log_2(freq^2 + 1)$ applied to these frequencies to flatten the distribution between the high and low frequency forms. This manipulation ensured the network would have sufficient opportunity to learn low frequency irregular forms. The final variance was between 1 and 21 tokens/inflected type, meaning that the most frequent words appeared just over twenty times as often as the least. We also carried out a separate series of simulations where a reduced frequency compression scheme was employed (a frequency variation of approximately 400:1 instead of 20:1). We report on the effect of this frequency compression manipulation below.

Simulations were run under two training schedules: mass training and incremental training. In mass training, all the stems and inflected forms are available to the network throughout the training process. In incremental training, training starts out with a small number of high frequency words. The training set is then gradually expanded to include words of decreasing frequency until the entire corpus is absorbed. These two training schedules are intended to capture the distinction between input (mass training) to the child and uptake (incremental training) by the child (Plunkett & Marchman, 1993). Incremental training is also designed to reflect the finding that highly frequent word forms are acquired early by children (Barrett, Harris, & Chasin, 1991; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Morrison, Chappell, & Ellis, 1997).

Analysis Techniques

All simulations were performed with five different random starting seeds to assess the consistency of the findings. Unless otherwise stated, all results reflect the mean over these five simulations. At the end of every epoch (which may contain over seventeen thousand

TABLE 1
Binary Encoding Based on Moby Pronunciator

Phoneme Code	Rough Pronunciation	Binary Vector
&	"a" in "dab"	0001001111000111
(	"a" in "air"	0010001011000111
@	"a" in "ado"	0010001111000011
A	"a" in "far"	0000001111000001
a	first half of "i" in "ice"	0001001111000001
b	"b" in "nab"	1110001000111111
C	"ch" in "ouch"	1110000000000011
D	"th" in "the"	1110001001001111
d	"d" in "pod"	1110001000000111
E	"e" in "red"	0010001111000111
e	first half of "a" in "day"	0011001111000111
f	"f" in "elf"	1110000001011111
g	"g" in "fig"	1110001000000001
h	"h" in "had"	1110000011000001
I	"i" in "hid"	0110001111000111
i	"e" in "see"	0111001111000111
J	"g" in "vegetable"	1110001000000011
j	"y" in "you"	0110001011000000
k	"c" in "act"	1110000000000001
l	"l" in "ail"	1110011011000111
m	"m" in "aim"	1110101000111111
N	"ng" in "bang"	11101010000001
n	"n" in "and"	1110101000000111
O	"o" in "dog"	0010001111000001
o	first half of "o" in "boat"	0011001111000001
p	"p" in "imp"	1110000000111111
R	"u" in "burn"	0010001011000011
r	"r" in "ire"	1110001011000111
S	"sh" in "she"	1110000001000011
s	"s" in "sip"	1110000001000111
T	"th" in "bath"	1110000001001111
t	"t" in "tap"	1110000000000111
U	"oo" in "book"	0110001111000001
u	"oo" in "too"	0111001111000001
v	"v" in "average"	1110001001011111
w	"w" in "win"	0110001011000001
Z	"s" in "vision"	1110001001000011
z	"z" in "zoo"	1110001001000111
#	silence/vacant	0000000000000000

individual pattern trials), the system was evaluated to determine what, and how much, had been learned. Every output pattern was examined to determine the nearest legal phoneme (using the RMS distance) in each template position, and then compared with the "correct" sequence. It should be noted that this is not the only possible way of identifying phonemes; an alternative method, where individual output units were threshholded to the closer of zero or one was also used in some analyses, but did not appear to produce markedly different results; these analyses have thus been omitted for brevity and clarity.

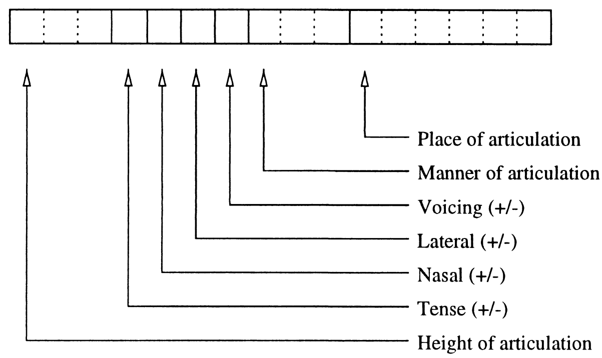

Figure 1. Composition of binary vector representation. Tenseness, Nasality, Laterality, and Voicing are binary features, coded as either on or off; Height, Manner, and Place are multivalued features and coded with a thermometer coding using a right-justified string of 1s.

Each output (type) from the network is analyzed as belonging to one of the following categories:

Correct. The network produced an inflected form identical to the training output for that form. Example: "cat/cats"

Regularized. The network produced an inflected form for a novel stem which was consonant with the "rules" for inflection (n.b. this is only applicable to novel forms such as nonce words or denominalized forms). Example: "wug/wugs" (pronounced /wugz/)

Over-regularized. The network produced an inflected form for an irregular training item which was partially consonant with the rules as above (n.b. this is only applicable to irregular forms). Example: "take/taked" or "take/tooked"

Irregularized. The network produced an inflected form similar to a known irregular form (n.b. this is applicable to both regular and irregular forms, as illustrated). Examples: "think/thunk" (similar to "sink/sunk") or "sight/sit" (similar to "light/lit")

Blend. The network produced an inflected form that shows evidence of partial irregularization as well as partial regularization (n.b. this is only applicable to regular forms, as otherwise these would be categorized as over-regularized). Example: "tow/tewed" (similar to "throw/threw" + -ed)

Wrongly Suffixed. The network affixed an incorrect suffix to an otherwise correctly inflected onset, nucleus, and coda. Example: "cat/catZ"

No Change. The network (incorrectly) reproduced the input stem exactly at the output. Example: "dish/dish"

Other. The network produced a form not otherwise classifiable above.

The careful definition of the above categories is important for comparative purposes, as a certain amount of ambiguity persists in the literature. For example, Marcus et al. (1992)

include inflected versions of past tense forms (e.g., "camed") as over-regularizations, while Plunkett and Marchman (1993) excluded them. This variation in definition makes direct numerical comparisons between various findings difficult. In this analysis, we have classified "camed" as partially (and incorrectly) regularized, and hence as an over-regularization error.

An important feature of the feed-forward networks employed in these simulations is that they are deterministic, in the sense that the same inputs (phonemes and syntax cues) will always produce the same outputs for a given weight set. Children, on the other hand, will sometimes switch forms within a single conversational turn (Kuczaj, 1977; Marcus et al., 1992). This makes direct numerical comparison of over-regularization rate somewhat problematic. We chose therefore to calculate error rates twice—on the basis of types and tokens. For example, over-regularization rates on the basis of tokens are given by the formula:

$$100\%\left(1 - \frac{\text{over-regularized tokens}}{(\text{over-regularized tokens} + \text{correctly inflected tokens})}\right)$$

It should be kept in mind, however, that since errors are more likely to occur on low frequency words, a type-wise assessment will yield a higher error rate than a token-wise assessment. Furthermore, there is considerable variation in error rates across children (see Marcus et al., 1992). We prefer to focus on the general profile of inflectional development rather than predict the profile for the individual child for whom, after all, we lack the relevant vocabulary and frequency data.

We tested generalization performance in a variety of ways. First, we compiled a corpus of cross-paradigmatic inflections; words (like "year") attested only in one syntactic category (in this case, as a noun). These words were presented to the network to be inflected in the other category, as a simulation of 1497 denominal verbs (to receive past tense) and 212 deverbal nouns (to be pluralized).[1] Furthermore, we collected 1541 "novel" forms, encompassing every noncapitalized (eliminating proper names, acronyms, etc.) monosyllable in the Moby corpus that did not duplicate, in spelling or pronunciation, an existing stem. We categorized the inflectional processes applied to these novel forms by the network. Finally, we examined the network's capacity to select the correct suffix by targeting different portions of the stem for presentation to the network. This provided an evaluation of the extent to which the network had abstracted a general representation of voicing assimilation and epenthesis.

III. SIMULATION 1

Introduction

Our first simulation involved presenting the entire training corpus to the network at each epoch (in random order) and updating the weights after each presentation. Our goal in this simulation is to establish a baseline performance for the *corpus* properties, as opposed to

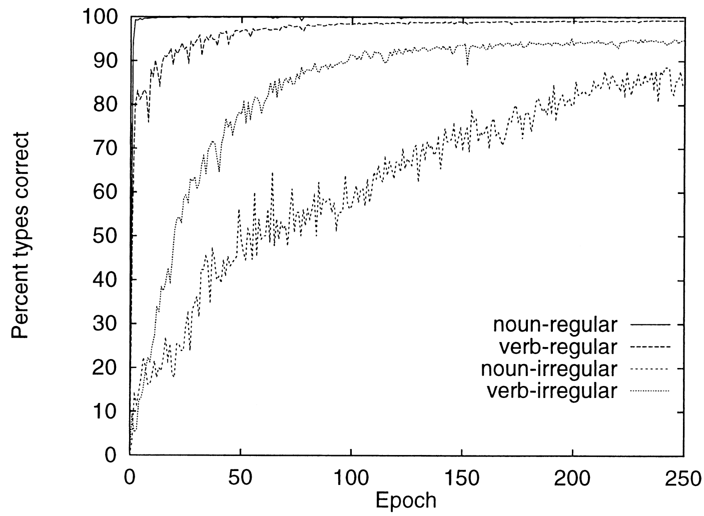

Figure 2. Percentage of types correct by training epoch (mass training).

the properties of the training schedule. We do not suppose that the child attempts to learn the inflections of all monosyllabic nouns and verbs simultaneously. Nevertheless, network performance under these training conditions provides a rough estimate of the relative ease of acquisition of the different inflectional types.

Results

Figure 2 shows that a single unified network is capable of learning both noun plural and verb past tense morphology; in fact, performance on the regular types is near-perfect after the first full epoch of training. Performance on irregular types lags significantly; performance after 200 epochs of training yielded 80% of irregular noun types and 95% of irregular verb types that were correctly inflected.

Discussion

After only a single epoch of training (which, recall, incorporated nearly 20,000 individual pattern trials), the system had mastered most of the basic "rules" for regular inflection of nouns. The dominant factor in both time course and eventual performance level for this sort of mass training is the sheer volume of a particular category that appears in the training set. The most common category, regular nouns, was learned near-perfectly and almost instantly relative to the other categories. On the other hand, irregular verbs are more common, both in terms of types and tokens (there are 997 irregular verb tokens (122 types) in the training set, compared with only 204 irregular noun tokens (26 types), and this is apparently enough to offset the fact that the category "irregular verb" appears to cover more widely varying and difficult inflectional paradigms than the category of irregular nouns. Another way of measuring this effect is by noting that irregular nouns constitute only 1.56% of the noun tokens and 1.14% of the noun types, while 24.4% and 12.90% of the verb tokens and types, respectively, are irregular.

In this simulation, individual token frequency does not play a decisive role; irregular forms are learned slowly and poorly, despite the fact that the individual tokens are among the most common in the entire training corpus. Specifically, we note a distinct absence of common, high-frequency forms such as “man/men” or “give/gave” being learned quickly relative to less frequent regular forms (such as “walk/walked”).

IV. SIMULATION 2

Introduction

An important consideration in the simulation of U-shaped learning curves is the initial acquisition of the correct forms for some irregular forms. Because the learning of a connectionist system is strongly influenced by the frequency of training patterns, rare forms are generally only learned well when they are made especially salient to such systems. The statistical dominance of regular forms tends to make these more influential unless irregular forms are somehow made more salient. In particular, the initial error-free period of irregular production can be problematic for connectionist systems (and does not, for instance, appear in Simulation 1).

The salience of the early acquired irregulars can be achieved by a manipulation of the training schedule to enforce learning of a small, usually irregular-rich, set of training samples which is then enlarged to a more representative final set. This manipulation can be justified in terms of children’s vocabulary development; although the words that children hear may be relatively constant, the words that children understand, and thus attend to, are a constantly increasing set that (presumably) starts with the more common words. We assume, then, that common words are more salient and we introduce an expanding corpus from more common to less common words. We thus followed standard practice (Elman, 1993; Plunkett & Marchman, 1993; 1996; Jackson, Constandse, & Cottrell, 1996) in our second simulation by beginning with a small (and irregular-rich) training set and gradually increasing the size of the training set at regular intervals.

The initial training set consisted of the twenty most frequent forms (comprising 375 tokens). Of these twenty forms, 12 were regular nouns, two were irregular nouns, five were irregular verbs, and one was a regular verb. The system was trained for five epochs, then the training set was increased by five percent (of current types) at every fifth epoch in decreasing token frequency order until the complete training corpus was presented at epoch 575 (115 incremental stages). Again, the system was run five times with different random seeds and the mean results are presented below in identical format.

Results

In analysis of incremental training, it is important to keep in mind the difference between performance on words already presented to (and known by) the network, and words that

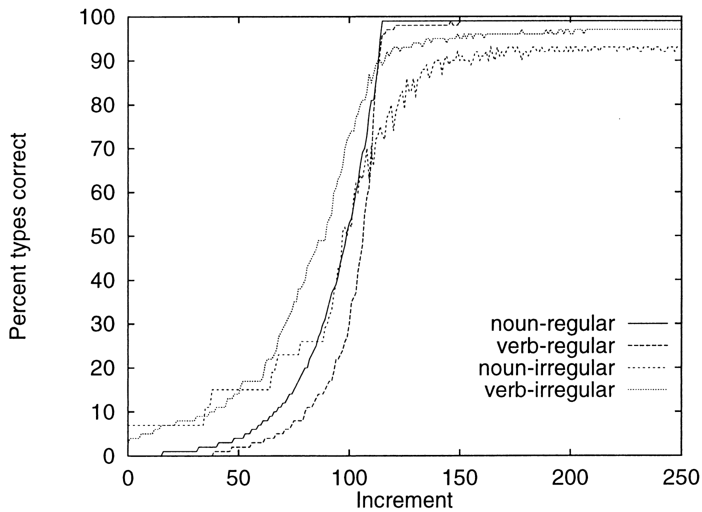

Figure 3. Percentage of types correct by training increment (incremental training). Note that incremental training reaches full vocabulary by increment 115. Training beyond this point corresponds to the mass training regime.

will eventually be part of the training corpus but have not yet been added. Accordingly two graphs (Figures 3 and 4) are presented, displaying the mean number of types correctly inflected as a function of the total corpus and as a function of the corpus thus far seen.

As training progresses, the network learns to inflect correctly the entire training corpus. Furthermore, we note the classic U-shaped curve in Figure 4. As in Simulation 1, we see

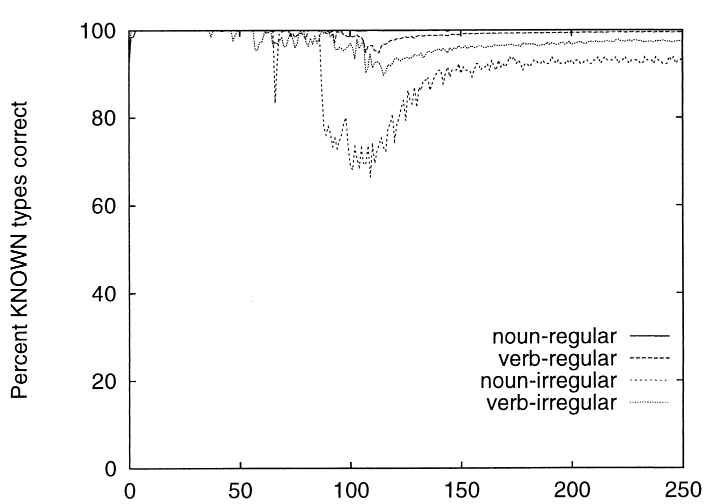

Figure 4. Percentage of *known* types correct by increment.

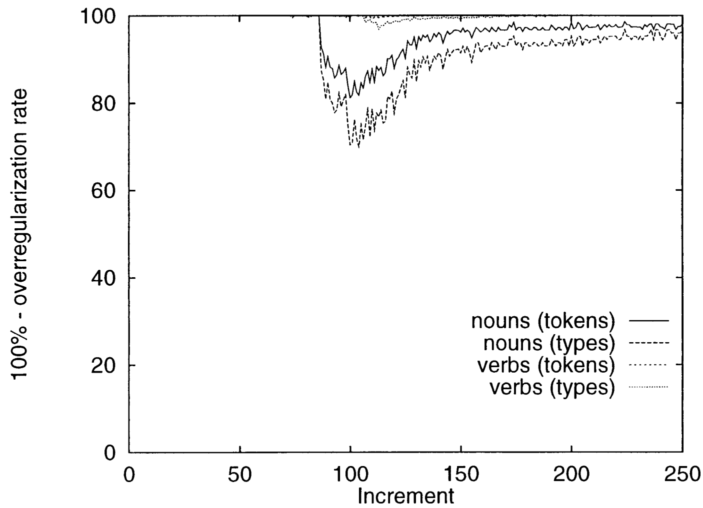

Figure 5. (100%-over-regularization rate) as a function of training increment (incremental training), until "increment" 116 at which the full corpus is reached, and increment becomes synonymous with "epoch."

that performance on regular nouns eventually dominates regular verbs, as well as irregulars of all sorts, and that irregular nouns are the most difficult overall. Regular nouns are also learned more quickly than regular verbs (Figure 3). However, Figure 3 clearly demonstrates an initial period where the networks' performances on irregular forms is superior to their performance on regular forms.

Final levels of performance in the two training regimes are remarkably consistent; the only substantial difference is a marginally lower error rate for irregular noun types in the second (incremental) experiment. However, the developmental course of the two experiments is radically different; the incremental network is able to master the noun and verb inflections irrespective of regularity until approximately the 90th training increment, at which point performance drops dramatically. This can be compared with Figure 5, where the over-regularization rate on nouns similarly grows explosively at the 90th increment, indicating a qualitative shift in representation despite a relatively continuous growth rate in vocabulary size. By the end of training, performance on noun plurals and verb past tense forms (types) is 99.9% and 99.2%, respectively.

Role of Frequency Compression. These exact performance figures, of course, are to some extent influenced by the composition of the training set. Varying, for instance, the amount of frequency compression/flattening will produce different performance levels. In particular, a less extreme flattening may cause infrequent forms (and paradigms) to be ignored in the sea of frequent forms. Common wisdom, which holds that irregulars are frequent and regulars are not, is not entirely correct in this assessment—for example the pairs "goose/geese", "sheaf/sheaves," and "flee/fled" are irregular, but have very low frequencies. The most common single category and paradigm is, by far, the regular noun.

To examine these effects, we ran additional simulations using a compression scheme of $\log_2(freq^2 + 1)^2$ (instead of $\log_2(freq^2 + 1)$), resulting in the most frequent type appearing about 400 times as often as the least (instead of 20). Under this scheme, the percentage of regular noun tokens is vastly *increased*, which increase is reflected in training. Using a similar incremental training regime, we found that regular nouns are still learned well (97.7% of types correctly inflected) but all other forms were significantly lost. Irregular nouns performed at only 76%, irregular verbs at 50%, and regular verbs at only 43%. Furthermore, most of the errors were miscategorization errors—the less frequent verb morphology was lost in an attempt to pluralize everything. This further confirms that salient but rare forms are difficult for a network to learn. These result also lend further credence to the view, originally suggested by Plunkett and Marchman (1991), that it is important to distinguish the *input* to the child from the *uptake* by the child. Raw frequency comparisons are unlikely to provide an accurate assessment of the relative saliency of individual word forms for the child. Here we have assumed that a high frequency compression captures the relative saliency of different word forms. Of course, a fuller account of the acquisition of inflectional morphology in children would need to take account of other factors, such as the word's meaning, in determining the saliency of a particular form.

Error Analysis

Over-regularization. The most widely studied measure of the acquisition of inflectional morphology is over-regularization rate. In accordance with other work (Marcus et al., 1992; Plunkett & Marchman, 1993; 1996), we define over-regularization rate as:

$$100\% \left(1 - \frac{\text{over-regularized tokens(types)}}{\text{over-regularized tokens(types)} + \text{correctly inflected tokens(types)}}\right)$$

The curves for both noun and verb types and tokens are presented in Figure 5. Every "increment" of the graph represents five complete passes through the corpus thus far, which includes from 20 (early training) to several thousand (late training) types. The maximum corpus size is achieved at increment 115, so increments 116–250 show the results of additional training on the complete corpus. Note that these curves do not represent the averages of 5 simulations. They are taken from an individual simulation. Average rates are reported below.

The network produces the characteristic developmental profile of a period of highly accurate production of a small, irregular-rich set of words, followed by a characteristic loss of accuracy on (some) irregulars as the system's experiential vocabulary increases and is more and more dominated by regular words. Performance then returns to near-perfect levels.

The mean over-regularization rate of the network for noun types was 11.50% $\pm$ 8.44% and 0.67% $\pm$ 1.25% for verb types. These averages were obtained by calculating means from the point of first observed over-regularization.

It is worth noting that the network starts to over-regularize nouns consistently earlier and continues to over-regularize nouns marginally longer than it over-regularizes verbs.

Studies of the time course and error rate of noun inflections in children are unfortunately rather rare; the best known is probably Marchman et al. (1997) and their response to the claims of Marcus (1995) about relative rates of over-regularization of nouns and verbs. The Marchman et al. (1997) study is limited in that they focused on the development of only five nouns and sixteen verbs, compared to the nearly 170 irregular forms modelled here. Nonetheless, they report an average over-regularization rate of 16% for nouns and 10% for verbs. Marcus (1995) reports comparable rates of over-regularization for both nouns and verbs of less than 10%. These results are broadly similar to those found in the network. Furthermore, Marchman et al. (1997) found that over-regularization of nouns happened both earlier and more frequently than those of verbs, again in line with the behavior of the network (and the predictions of Marcus, 1995).

Categorial Error Analysis. Tables 2 and 3 present an analysis of the mean number of overall error types. The first sixty incremental stages are omitted as being near-perfect. The tables' extension past increment 115 represents additional training (five epochs per "increment") with the final and complete vocabulary set. This tabular format was chosen because of the difficulties in establishing reasonable base-line corpus comparisons as well as the fact that the nature of network errors changes over time. As a general comparison, however, the mean over-regularization rate for noun types (of all samples) was 7.46%, while all other categories had less than 0.01% error rate; for verb types, the mean over-regularization rate was 0.45% which is of the same order of magnitude as suffixation errors (0.12%), no-change errors (0.18%), and unanalyzable errors (0.39%), reflecting a wider variety of error categories (all other verb categories were below 0.01% of types).

Marcus et al. (1992) state that children make irregularization errors significantly less frequently (i.e., an order of magnitude less) than over-regularization errors. It is worth noting, then, that even using the broadest possible definition of "irregularization" (including what we have here categorized as "blend") at the worst possible time for the network (at 110 increments), we calculate an irregularization rate of only 0.67% for verb types (and significantly less for noun types), compared with the much higher over-regularization rate of 3.68% for irregular verbs.

A significant finding is that the most frequent irregular words were remarkably resistent to over-regularization; no word with a token frequency of greater than 243 (15, in our training set) was ever over-regularized. Thus, moderately common words like "keep," "tell," and "let" (as well as extremely common forms like "see" and "man") were immune to over-regularization, while only marginally less common words like "wife," "child," and "hold" were over-regularized upon occasion.

Phonological conditioning of errors was also apparent on the verb forms. Although few in number throughout training, "no change" errors were more likely to occur on stems that end with an alveolar consonant. Furthermore, "no change" verbs (like hit/hit) were less likely to be over-regularized than other irregular verbs. Of the 122 irregular verb types, 27 (22.13%) are "no change" verbs. Of the 18 types that are ever attested as being over-

TABLE 2
Mean Number of Noun Type Errors of Each Category under Study (Incremental Training)

	Overreg	Blend	Irreg	Suffix	N. C.	Other	Irregs	Regs
60	0.0	0.0	0.0	0.0	0.0	0.0	4	177
65	0.0	0.0	0.0	0.0	0.0	.6	5	220
70	0.0	0.0	0.0	0.0	0.0	0.0	6	278
75	0.0	0.0	0.0	0.0	0.0	0.0	6	351
80	0.0	0.0	0.0	0.0	0.0	0.0	7	453
85	0.0	0.0	0.0	0.0	0.0	0.0	7	579
90	1.4	0.0	0.0	0.0	.8	.4	10	750
95	3.6	.2	0.0	0.0	0.0	.2	15	944
100	6.2	0.0	0.0	0.0	0.0	0.0	20	1198
105	6.2	.2	0.0	0.0	0.0	2.4	23	1495
110	6.6	.4	0.0	0.0	0.0	.4	25	1827
115	7.0	1.2	0.0	0.0	0.0	2.2	26	2254
120	6.2	.6	0.0	0.0	.4	1.4	26	2254
125	5.2	.6	0.0	0.0	.2	1.4	26	2254
130	4.4	.6	0.0	0.0	.2	1.0	26	2254
135	2.8	.6	0.0	0.0	0.0	1.0	26	2254
140	2.4	.6	0.0	0.0	0.0	1.0	26	2254
145	2.6	.6	0.0	0.0	0.0	1.2	26	2254
150	2.6	.4	0.0	0.0	0.0	1.0	26	2254
155	2.8	.4	0.0	0.0	.2	1.0	26	2254
160	2.2	.2	0.0	0.0	0.0	1.0	26	2254
165	2.2	.2	0.0	0.0	0.0	1.0	26	2254
170	2.0	.2	0.0	0.0	.2	1.0	26	2254
175	1.8	.2	0.0	0.0	0.0	1.2	26	2254
180	1.8	.4	0.0	.2	.0	1.2	26	2254
185	1.8	.2	0.0	0.0	0.0	1.2	26	2254
190	1.8	.2	0.0	0.0	0.0	.8	26	2254
195	1.8	.2	0.0	.2	.0	.8	26	2254
200	1.6	.2	0.0	.4	.2	.4	26	2254
205	1.8	.2	0.0	0.0	0.0	.4	26	2254
210	1.6	.2	0.0	.4	.2	.4	26	2254
215	1.4	.2	0.0	0.0	.4	.4	26	2254
220	1.4	.2	0.0	0.0	.4	.4	26	2254
225	1.6	.2	0.0	.2	.2	.4	26	2254
230	1.6	.2	0.0	0.0	.2	.2	26	2254
240	1.4	.2	0.0	0.0	.2	.4	26	2254
250	1.2	.2	0.0	0.0	.4	.4	26	2254

regularized, two (11.11%) are "no change" verbs. No change verbs are thus overregularized at half the expected rate.

Generalization

We also tested the model on its ability to generalize to novel forms as training progresses—the so-called "wug" test. This involves evaluating the extent to which the network produces the correct inflection on a novel stem, which in this context means the form

TABLE 3
Mean Number of Verb Type Errors of Each Category under Study (Incremental Training).

	Overreg	Blend	Irreg	Suffix	N. C.	Other	Irregs	Regs
60	0.0	0.0	0.0	0.0	0.0	.6	22	28
65	0.0	0.0	0.0	0.0	1.0	0.0	29	38
70	0.0	0.0	0.0	0.0	.6	.8	37	49
75	.8	0.0	0.0	1.0	.2	1.0	45	68
80	0.0	0.0	0.0	1.0	0.0	2.2	51	86
85	0.0	0.0	0.0	0.0	.4	1.4	60	113
90	0.0	0.0	0.0	.8	0.0	.4	65	139
95	1.0	0.0	.2	.2	.2	2.4	81	188
100	1.2	1.2	0.0	1.0	0.0	3.2	92	255
105	1.2	.4	.4	1.6	.6	5.0	101	377
110	4.2	2.6	2.0	4.0	3.6	12.6	114	578
115	5.2	3.8	2.4	5.0	4.6	14.6	122	824
120	2.8	3.0	1.6	3.8	3.2	9.4	122	824
125	2.4	1.8	2.0	1.6	3.0	6.4	122	824
130	1.6	2.2	1.2	1.6	2.8	6.4	122	824
135	.8	2.0	1.6	1.4	2.8	4.8	122	824
140	.6	1.8	1.2	1.4	2.4	4.4	122	824
145	.6	1.2	1.0	.2	2.2	6.2	122	824
150	1.0	.8	1.0	.6	2.4	3.6	122	824
155	.6	.8	1.0	.2	2.4	3.0	122	824
160	.2	1.0	1.0	.2	2.8	2.8	122	824
165	.2	1.0	1.4	0.0	2.4	2.0	122	824
170	.2	.8	1.0	.2	2.4	2.4	122	824
175	.6	.8	1.2	.2	2.4	1.6	122	824
180	.2	.8	1.2	.2	2.4	1.6	122	824
185	0.0	1.0	1.0	.2	2.6	1.8	122	824
190	0.0	.6	1.2	.2	2.2	2.0	122	824
195	0.0	.4	1.0	.2	2.0	1.6	122	824
200	.2	.4	.8	0.0	2.2	2.4	122	824
205	0.0	.2	1.2	0.0	2.0	1.4	122	824
210	.2	.4	1.0	1.0	1.8	1.2	122	824
215	.2	.2	1.2	.6	1.6	1.4	122	824
220	0.0	.2	1.0	0.0	1.6	1.0	122	824
225	.2	.6	1.0	0.0	1.2	1.2	122	824
230	0.0	.2	1.0	0.0	1.6	1.0	122	824
240	0.0	.2	1.0	0.0	1.6	1.0	122	824
250	.2	.2	1.0	.2	1.4	1.0	122	824

consonant with the rule-based description. For this test, we used 1541 monosyllables taken from the Moby pronunciator that were not part of the Kučera-Francis data. The developmental course of generalization to novel stems is depicted in Figure 6. For purposes of comparison, we also include the profile of generalization to novel forms under the mass training schedule used in Simulation 1.

Figure 6 shows that the network is able to generalize the correct suffix to novel nouns and verbs. By the end of training over 90% of all novel nouns are inflected with the correct suffix as are over 80% of all novel verbs. Secondly, the final level of generalization of the

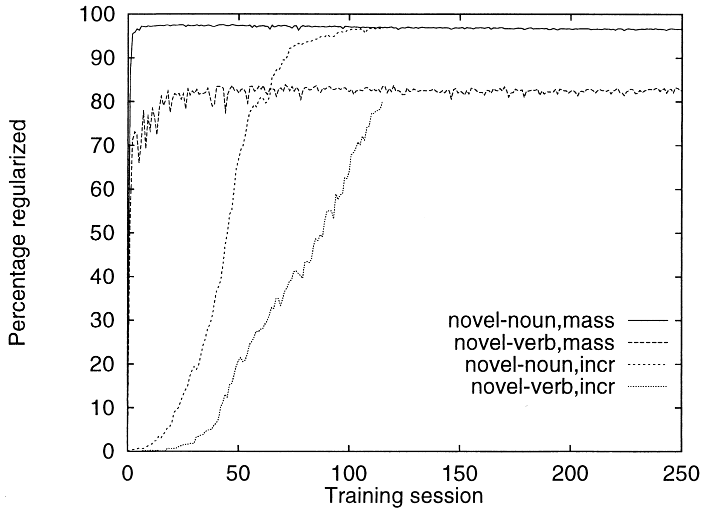

Figure 6. Regularization of nonce words by epoch or increment across training regimes.

network is independent of the training regime adopted (mass versus incremental) indicating that generalization is determined more by the training corpus than by the training method.

Generalization also shows a pronounced developmental profile for the incremental training regime. During early training, the network is unable to add an appropriate suffix to a novel form. As the training vocabulary expands the generalization performance increases exponentially until it asymptotes at the levels indicated above. However, the developmental profile of generalization differs for nouns and verbs. The network regularizes novel verbs considerably later than novel nouns.

We also analyzed the likelihood of appending a suffix to a novel verb contingent upon its form. In English, most irregular verbs that take "null suffixes" have stems that end in alveolar stop consonants (/d/ or /t/), such as "light/lit", "hit/hit", and "sit/sat". Pairs ending in velar consonants ("dig/dug", "stick/stuck") are rarer, and pairs ending in dental fricatives are rarer still; in the training corpus we used, only one stem/inflection pair ("tooth/teeth") ended in a dental fricative, and that, of course, is a noun inflection and not a verb. We examined the percentage of nonce verbal forms that were inflected regularly (i.e., correctly suffized) against those that were inflected without a suffix. Table 4 shows the distribution of null suffixes over selected stem endings.

Table 4 indicates that a higher percentage of novel verbs ending in alveolar stops (such as /t/) receive "null suffixes" than those ending in dental fricatives. In particular, four times as many nonce forms ending in dental fricatives were regularized as were null-inflected, while this ratio was nearly reversed for nonce forms ending in alveolar stops.

TABLE 4
Distribution of Null Suffixes over Selected Novel Verb Stem Endings

Category	% Regularized	% "Null" Inflected	Ratio (R/N)
Verb, alveolar stop (d/t)	12.7 (%)	39.9 (%)	0.318
Verb, velar stop (g/k)	43.1	30.1	1.332
Verb, labial stop (b/p)	38.4	20.9	1.837
Verb, dental fric. (D/T)	45.5	9.52	4.779

Critical Mass Effects

The profiles of development in the incremental training schedule suggest that there is a close relationship between levels of generalization and performance, on the one hand, and the number of words in the training set, on the other. For example, the profiles of development depicted in Figure 6 suggest that a critical mass of nouns and verbs is required in the training set before high levels of generalization are achieved. Similarly, the sudden onset of over-regularization errors depicted in Figure 5 indicate a mass action effect at work. These findings are consistent with earlier work (Plunkett & Marchman, 1993; Marchman & Bates, 1994). However, in the current simulations there is substantial delay between sudden increases in the various performance measures for nouns and verbs. The critical mass hypothesis predicts that these delays are directly related to the number of nouns and verbs in the corpus at different points in training.

Figure 7 shows the relationship between the average number of novel forms regularized and the number of types of a given syntactic category present in the training set at each point.

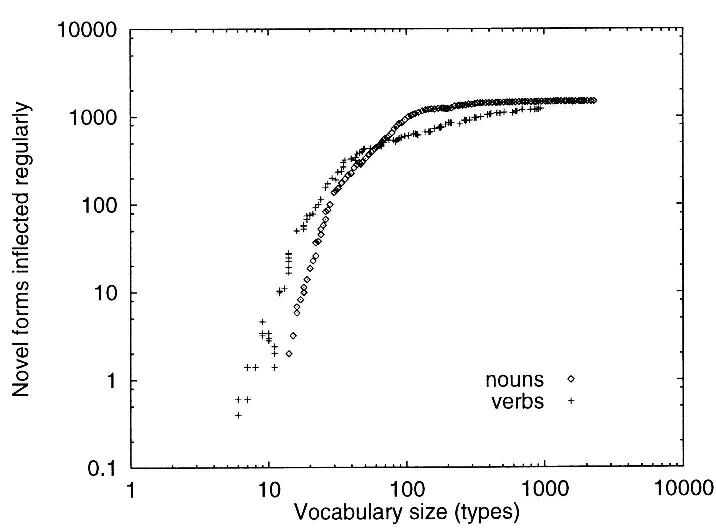

Figure 7. Evidence for critical mass effect; regularization rate against type rate scatter plot.

Figure 7 (in conjunction with Figure 6) provides support for this critical mass hypothesis. This graph plots the regularization rate for novel forms against the number of types *of the same category* in the current training increment. The curves for nouns and verbs in Figure 7 are nearly identical, showing that a similar process of mass action is operating for both syntactic types. Both curves show a rapid increase in regularization rate, beginning to saturate at about 100 types of the appropriate syntactic category. The *developmental* delay between nouns and verbs shown in Figure 6 reflects the composition of the input vocabulary at different stages in training. The small advantage of nouns over verbs in Figure 7 reflects the reflective homogeneity of noun inflections compared to verbs (recall that there are many more irregular verbs than irregular nouns).

Assimilation Effects

One of the more prominent aspects of noun and verb inflection, studied jointly, is that the voicing of the final phoneme of the stem is critical in determining the correct inflection for either a plural or past tense. Both inflectional types involve perseverative assimilation, where a feature of a prior sound persists and attaches itself additionally to another sound. In the human vocal tract, this assimilation can be explained by the observation that changing voicing requires time and energy and can thus be lost in fast, fluent speech, resulting in the stem-final voicing feature causally affecting the voicing of the suffix.

The argument and architecture presented by Pinker and Prince (1988) attempts to capture this in a separate voicing module, which is “down-stream” from the stem-affixation module and examines (and projects) the voicing appropriately. Pinker and Prince (1988) argue that the Rumelhart and McClelland (1986) model is unable to capture the redundancy of morphophonological and phonological processes which apply across paradigms. They discuss at some length the virtue of rule systems, and argue that the main contribution of the rule approach is *not* (as they claim Rumelhart and McClelland (1986) suggest) to account for U-shaped phenomena. Rather, it is the ability of rules to factor out common processes which cut across different domains. Voicing assimilation and epenthesis are two rules which they give as examples, applying not only to the past tense, but also to 3rd person singular, noun plurals, the possessive, contractions, etc. This analysis penetrates to the heart of what generative linguists like about rule systems: the notion that a handful of rules may recurr, interact, and generate a fair amount of apparent complexity. There is enormous combinatoric power in rule systems. Pinker and Prince (1988) criticized the Rumelhart and McClelland (1986) model for implying that processes like epenthesis and voicing assimilation are idiosyncratic phenomena which are peculiar to the past tense paradigm. They claim, strongly, that there is no way for this information to be shared across different corners of the grammar. We argue that the modular decomposition demanded by Pinker and Prince (1988) is not necessary and that our network is capable of capturing such regularity within a single system.

The causal structure of a neural network is somewhat different from the physics of a vocal tract; as each unit is independent and unordered (within a layer), there is by construction no *causal* link between independent output units. There is, however, an *associative* link between independent outputs—and a possible causal link between the voicing feature of the final consonant of the *input* stem and the voicing feature of the final consonant of the (output) suffix.

To investigate this possibility, “don’t care” inputs (units with output activations of 0.5) were used to construct a 128-bit “neutral” stem. Each feature/bit in the final phoneme was set, then reset, and presented to the five incrementally trained networks at their final weight states to inflect as a noun, a verb, and as a “don’t care” inflection. All other bits in the final phoneme remained at “don’t care”; so the test vector consisted of only one meaningful bit. The differences between output activations with the feature set and unset were evaluated as a measure of the significance of that feature. If a strong causal link existed between input voicing and output voicing, then toggling the voicing bit in the input should produce an output with a strongly varying voicing.

No evidence for this level of learned voicing assimilation, however, was found. In point of fact, no single feature change produced a substantive change in the output representations of the final consonant. Major effects were found of some features on the production (or lack thereof) of an epenthetic vowel—the features by which an epenthetic schwa differs from an empty slot were strongly affected by some features (such as nasality, or labialism) that strongly predicted the absence of such an epenthetic vowel.

This finding should not, however, be interpreted as demonstrating that these feed-forward networks—or neural networks in general[2]—cannot acquire “principles” such as voicing assimilation. The evidence from the generalization tests, presented above, clearly indicates that these networks have mastered voicing assimilation at the level of the entire, pronounceable word. We thus conclude, not that voicing assimilation is impossible for such networks, but that voicing assimilation (and possibly other properties) can be described in a subsymbolic way as a property of the representation of an entire word.

To investigate this possibility further, we tested network performance on novel stems presented with “don’t care” syntactic units. Despite the absence of meaningful inflection instructions, the networks still usually produced meaningful and recognizable attempts at inflection. Unexpectedly, despite the tremendous dominance of both noun types and noun tokens, novel stems were inflected as nouns and verbs with nearly equal frequency. Finally, and most importantly for a study of inflectional assimilation, the number of voicing errors was small; of 7705 novel stem trials (1541 novel stems times 5 different “subjects”), only 927 were “inflected” to produce a suffix-final consonant with different voicing than the stem-final consonant, and of those 927 “errors”, 216 were examples of the t/@d pair (as in “swat/swatted”), which is, of course, correct in context. We thus observe a maximum of 9.2% voicing errors, even in the absence of syntactic information, clear evidence that these networks can successfully generalize to produce principles such as cross-inflectional voicing assimilation.

Denominal and Deverbal Forms

Recent work by Kim et al. (1991), Kim, Marcus, Pinker, Hollander, and Coppola (1994) has demonstrated that speakers do not rely exclusively on phonological information to

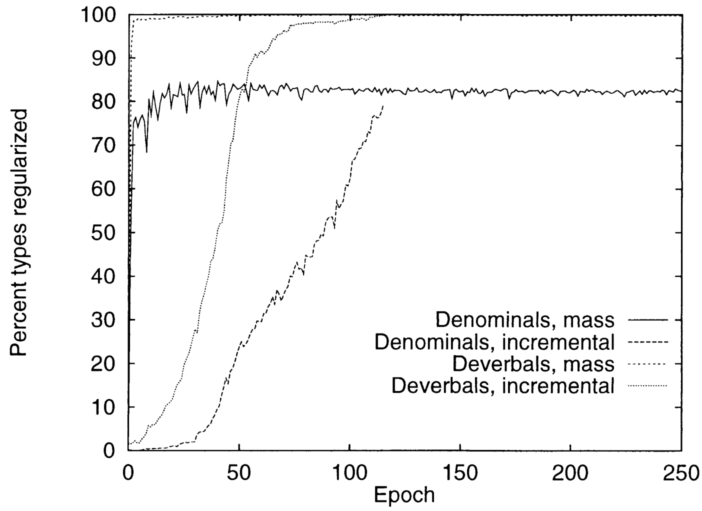

Figure 8. Network performance on cross-paradigmatic inflections.

determine the manner in which verb forms are inflected. For example, a noun stem which is identical to the stem of an irregular verb will nevertheless be regularized when used as a denominal verb. Unfortunately, it is not possible in the present simulations to model this process since the network, by design, only has access to phonological information. Nevertheless, it is possible to evaluate the extent to which the network regularizes noun stems and verb stems when they are inflected outside of their trained categories. Therefore, we analyzed the performance of the network on cross-categorial (denominal and deverbal) forms by presenting the network with all of the unicategorial forms in the wrong category. This gave us a set of 1497 denominal verbs and 212 deverbal nouns to be inflected in the new category. Figure 8 indicates the percentage of regularized types as a function of training epoch. Again, for comparison, we also plot the same measure for the mass training schedule reported in Simulation 1.

By the end of training over 80% of denominal verbs and over 90% of deverbal nouns are regularized. These results indicate that the network clearly has the ability to produce cross-categorial (denominal or deverbal) inflections despite having no direct (training) evidence about their status as cross-categorial stems. Note also that the regularization profiles for cross-paradigmatic inflections (Figure 8) and generalization to novel stems (Figure 6) are almost identical. The cross-paradigmatic inflections are treated as though they are novel forms. In other words, information about the derivational history of these unicategorial stems is not a causal factor in determining their high level of regularization.

A slight modification of the network permits a more direct comparison of the Kim et al. (1991, 1994) results. We enlarged the input layer of the network moderately, adding to each input type a random set of 50 bits as a form of random pseudo-semantics. These semantics were chosen by simply setting (or not) each bit independently with a 50% probability. The network was then trained to associate the appropriate inflected form with the stem, the syntactic category, and the "semantics" appropriate to that form. This constitutes training the network to inflect various forms in the context of a representation of their conventional meanings.

We then tested the performance of the system (after 115 increments) in producing inflected forms of irregular stems associated with new and previously unseen "semantics," reflecting a stem used in an atypical or novel context, such as with a completely new denominalized form. These new semantics, which can formally be regarded (in this model) as high-level noise, will of course distort the outputs in a way similar to lesioning. We expect, therefore, a certain level of degraded performance and simple errors. The important question is whether the analyzable forms produced reflect the original irregular inflection, or are regularized in keeping with human practice on novel meanings and form.

The data from Kim et al. (1991) suggest that most of these stems should be regularized; we find that with 50 units of pseudo-semantics, 11 (of 26) noun stems are regularized, compared with only three with the "correct" irregular inflection. Similarly, of 122 verb stems, we find 32 stems regularized and only seven irregularized. In either case, the number of regularized forms exceeds the number of irregular forms by about 4:1, a clear demonstration of the ability of the network to produce "regularized" forms when presented with novel uses and meanings. Varying the number of semantic units shows that as the number of such units increases, and the importance of semantics to the overall representation of the word-form correspondingly increases, the number of stems regularized also increases. These results suggest that human-like levels of denominal regularization can be obtained by using an appropriate number of semantics units.

V. DISCUSSION

The simulations reported here track the developmental trajectory of a neural network trained on a realistic corpus of English nouns (2280 types) and verbs (946 types). Indeed, the model incorporated *all* the monosyllabic nouns and verbs from the Brown corpus.[3]

The performance of the networks mimics that of children and adults in a number of important respects. First, all simulations (mass and incremental) are able to learn the training corpus to near-perfection. Recall that by the end of training, performance on noun plurals and verb past tense forms is 99.9% and 99.2%, respectively. Second, the profile of over-regularization errors for both nouns and verbs in the network mimics the well-documented U-shaped profile of development in children. For verb past tense forms, this result replicates and extends the findings reported for models trained on smaller vocabularies (Plunkett & Marchman, 1993, 1996). However, this is the first demonstration in which a neural network model has simulated the well-known U-shaped development for noun plurals, in particular producing a initial phase of error-free performance. This finding shows that Marcus' (1995) criticism of the original Rumelhart and McClelland (1986) model is unwarranted:

> Because irregular noun plurals are so rare, there is unlikely to ever be a stage in which irregular plurals dominate regular plurals; hence the Rumelhart & McClelland model would probably over-regularize even its earliest plurals. (p. 450)

Our model not only produces initial error-free performance on noun plurals but does so in the context of initial error-free performance on verb past tense forms.

The onset of over-regularization errors on nouns tends to occur earlier than over-regularization errors on verbs in the network. This result is consistent with the data reported in Marchman et al. (1997). The data for the four children analyzed in Marcus (1995) are heterogenous in this respect—Adam over-regularizes nouns before he over-regularizes verbs, Eve and Sarah show the reverse pattern and Abe starts over-regularizing nouns and verbs around the same time.[4] However, both studies demonstrate that the rate of over-regularization for nouns is greater than that for verbs. This is also true of the simulation as can be seen in Figure 5. Children make few over-regularization errors on forms with a high token frequency (Marcus et al., 1992). Again, this is true of the current set of simulations, replicating the results of earlier work with smaller verb vocabularies and demonstrating that the frequency effect scales up to larger vocabularies and extends to noun morphology.

The model also mimics children's early acquisition of regular noun plurals relative to regular past tense forms reported in previous empirical investigations (Brown, 1973; Marchman et al., 1997). This result is apparent from the network's performance on the training data (see Figure 3) and its ability to generalize to novel forms (see Figure 6). We do not suppose that the earlier acquisition of regular noun plurals in the network provides a complete explanation of why regular past tense forms should be acquired later in children. Presumably, conceptual, semantic and grammatical factors also have a role to play in this acquisition story. However, our results indicate that phonological and frequency factors (the only sources of information available to the network) may contribute much to the variance in the acquisition rate for these two grammatical forms in young children.

Regular nouns are always easier to learn than regular verbs for the network. However, this pattern does not hold for irregular nouns and verbs. During the earliest stage of training, irregular nouns have the advantage over irregular verbs. As training proceeds performance on irregular verbs consistently exceeds that of irregular nouns (see Figures 2, 3, and 4). We know of no detailed empirical investigation that charts the acquisition of irregular noun plurals relative to irregular verb past tenses in young children. Brown (1973) indicates the early acquisition of some irregular plural forms. The prediction of the model that the class of irregular past tense forms subsequently becomes easier to acquire than the class of irregular plurals (despite the former being more numerous) must await further investigation. From the point of view of network learning, this regularity by word category interaction is unsurprising; regular nouns, the most frequent type of inflection, were learned fastest and most accurately, while the network had the most difficulty with irregular nouns, the rarest category. The critical mass analysis also shows that the number of nouns/verbs in the training set is the factor determining generalisation to novel forms (Figure 7). The networks show high levels of generalization (appending the appropriate suffix ending) to novel stems only when they have learned around 100 forms from the relevant category (nouns or verbs). Marchman and Bates (1994) report parallel findings for young children.

Errors in the simulations show evidence of phonological conditioning. Networks make a small number of "no change" errors on verbs throughout training (Table 3). These are more likely to occur on stems that end with an alveolar consonant (table 4), e.g., *tread* → *tread* and *rend* → *rend*. This pattern of response is consonant with that reported for children by Bybee and Slobin (1982). The networks are *less* likely to make over-regularization errors on "no change" irregular verbs. In fact, the rate of over-regularization of other irregular verbs is double that of the "no change" sub-class. This result parallels the finding of Marcus et al. (1992) that children's "no change" past tense forms are resistant to over-regularization. The networks make very few no change errors on nouns (Table 2). We know of no empirical reports on "no change" errors for nouns in children. As can be seen from Table 3, irregularization is the least frequent error type produced on nouns and verbs. After over-regularization errors, the most common error type is "no change" then blend. "No change" errors are the predominant verb error type during the later phases of training. A similar rank ordering of error types is observed for nouns though the absolute level of errors is less than that observed for verbs. Again, the pattern of errors on verbs is consonant with that reported for past tense errors in children (Kuczaj, 1977; Marcus et al., 1992). Lack of empirical data prevents us from evaluating the rank ordering of error types for nouns.

Performance on denominal verbs and deverbal nouns indicates that cross-categorial generalisation is not difficult for networks of this type. The network has no difficulty learning to inflect two phonologically identical forms in quite different ways when they are taken from different syntactic classes. In particular, a stem which behaves as an irregular noun plural can behave as a regular verb past tense form within this system (or *vice versa*). For example, the irregular noun plural *men* is treated as a regular verb in its past tense form *manned*. These networks cannot (by design) add a regular verb inflection to a noun stem which is identical to the stem of an irregular verb (it only has access to phonological information). However, the tendency of these networks to regularize across categories suggests that the provision of other information such as meaning and/or distributional properties might well result in the regularization of noun stems that are phonologically identical with an irregular verb when used as a verb. This was confirmed in a set of supplementary simulations in which pseudo-semantic vectors were used to distinguish novel usage of irregular verbs from their routine usage. Well-formed phonological outputs to irregular stems were predominantly regularized. Similar results held for deverbals that were identical in phonology (but not semantics) to irregular nouns.

Finally, the capacity of these networks to choose the correct inflectional allomorph to attach to a novel stem demonstrates that processes controlling voicing assimilation can be observed in these networks. This result is obtained even in the absence of syntactic information. Clearly, the network exploits information about stem-final voicing even at the same time as it attempts to determine whether an input should be inflected as a regular or irregular form.

VI. CONCLUSION

The dual-route model of inflectional morphology, as proposed by Pinker and Prince (1988), has the advantages of being easy to understand and to decompose, and has an easily described and tested generalization performance. Unfortunately, like many symbolic descriptions of human mental processing, the simple and clean solution to a small problem (verb past tenses) becomes significantly less simple and clean when scaled up (is yet another route required for noun plurals?) and can result in needless complexity. Furthermore, the many similarities between these two processes of noun and verb inflection argue for a closer coupling than is required by the dual-route model. The complex requirements of the similarities and differences in processing argue for an equally complex, and unparsimonious, modular decomposition of thought, seeded through with questions about module ordering, data flow, and so forth.

We have presented a single-route or, more accurately, single-process model, based on a connectionist associative network, that is capable of inflecting verb stems to produce their past tense forms or noun stems to produce their plurals. It handles both regular and irregular verbs with reasonable accuracy, despite having a much larger vocabulary than many related projects e.g., Daugherty and Seidenberg (1992), Plunkett and Marchman (1991, 1993). Furthermore, it produces linguistically plausible generalizations, capturing important aspects such as performance on nonce words, the generalization of voicing assimilation even in the absence of syntactic information, and regularized inflection of cross-categorial processes such as denominal verbs or deverbal nouns. Importantly, the model mimics well-established facts about children's acquisition of noun and plural morphology, as well as offering several novel empirical predictions.

This system thus demonstrates that a single route to inflectional morphology is capable of producing the generalization and productivity levels necessary for a psycholinguistically meaningful model of noun and verb inflection, even on a very limited set of information (excluding semantics entirely, for example) and provides a baseline performance describing what can be done in this limited domain and to which new performance can be compared, as representations and information improve.

Acknowledgments: We thank Jeff Elman, Mary Hare, and two anonymous reviewers for comments on earlier drafts of this manuscript. This work was supported by a research project grant from the Economic and Social Reasearch Council to KP.

NOTES

1. An additional set of simulations dealing with the case of denominal verbs which are homophonic with extant irregular verbs is described in a later section.
2. It is worth noting that assimilative and epenthetic processes have already been demonstrated in other types of network architectures such as recurrent networks. See Hare (1990), Gasser and Lee (1992) and Cottrell and Plunkett (1994) for examples of recurrent networks that deal with assimilative processes within single inflectional paradigms. However, our aim if to evaluate whether these processes can be shared across inflectional paradigms when there are only associative connections rather than casual links between consecutive output sounds.

3. MacWhinney and Leinbach (1991) describe a simulation incorporating an even larger corpus of verbs and a wider range of inflections. However, their paper does not provide a detailed analysis of the developmental trajectory of the network's performance. In particular, they provide no account of early U-shaped learning.
4. Onsets of verb over-regularization for these four children are taken from Marcus et al. (1992).

REFERENCES

Baayan, H., Piepenbrock, R., & Rijn, H. van. (1993). *The CELEX lexical database (CD-ROM)*. University of Pennsylvania, Philadelphia: Linguistic Data Consortium.

Barrett, M., Harris, M., & Chasin, J. (1991). Early lexical development and maternal speech: A comparison of children's initial and subsequent uses of words. *Journal of Child Language, 18*, 21–40.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Bybee, J., & Slobin, D. (1982). Rules and schemes in the development and use of the English past tense. *Language, 58*, 265–289.

Christiansen, M. H., & Chater, N. (1999). Connectionist natural language processing: The state of the art. *Cognitive Science, 23*, 000–000.

Cottrell, G., & Plunkett, K. (1994). Acquiring the mapping from meaning to sound. *Connection Science, 6*(4), 379–412.

Daugherty, K., & Seidenberg, M. S. (1992). Rules or connections? The past tense revisited. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (p. 259–264). Hillsdale, NJ: Erlbaum.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71–99.

Ervin, S. M. (1964). Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), *New directions in the study of language* (pp. 163–189). Cambridge, MA: MIT Press.

Fromkin, V. A. (Ed.). (1973). *Speech errors in linguistic evidence*. The Hague: Mouton.

Garrett, M. F. (1980). The limits of accommodation. Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, hand* (pp. 263–71). New York: Academic Press.

Hare, M. (1990). The role of similarity in Hungarian vowel harmony: A connectionist account. *Connection Science, 2*(1), 123–150.

Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition, 56*, 61–98.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236–248.

Jackson, D., Constandse, R. M., & Cottrell, G. W. (1996). Selective attention in the acquisition of the past tense. In *Proceedings of the Eighteenth Annual conference of the Cognitive Science Society* (p. 183–188). Mahwah, NJ: Erlbaum.

Juola, P., & Zimmermann, P. (1996). Whole-word phonetic distances and the PGPfone alphabet. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP-96)*. Philadelphia, PA.

Kim, J., Marcus, G., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15*, 173–218.

Kim, J. J., Marcus, G. F., Pinker, S., Hollander, M., & Coppola, M. (1994). Sensitivity of children's inflection to grammatical structure. *Journal of Child Language, 21*(1), 173–210.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16*, 589–600.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualisations: Revising the verb learning model. *Cognition, 29*, 121–157.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 331–6.

Marchman, V., Plunkett, K., & Goodman, J. (1997). Over-regularization in English plural and past tense inflectional morphology. *Journal of Child Language, 24*(3), 767–779.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*(3), 189–256.

Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T., & Xu, F. (1992). Over-regularisation in language acquisition. *Monographs of the Society for Research in Child Development, 57*(Serial No. 228).

Marcus, G. F. (1995). Children's over-regularization of English plurals: A quantitative analysis. *Journal of Child Language, 22* (2), 440–60.

Miyata, Y. (1991). *A user's guide to PlaNet version 5.6: A tool for constructing, running, and looking into a PDP network* (electronic software manual).

Morrison, C., Chappell, T., & Ellis, A. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology: Section A—Human Experimental Psychology, 50*, 528–559.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73–193.

Pinker, S., & Prince, A. (1991). Regular and irregular morphology and the psychological status of rules of grammar. In L. Sutton (Ed.), *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, California: Berkeley Linguistics Society.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38*, 43–102.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.

Plunkett, K., & Marchman, V. A. (1996). Learning from a connectionist model of the English past tense. *Cognition, 61*, 299–308.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Parallel distributed processing: Explorations in the micro-structure of cognition. In Psychological and Biological Models, (Vol. 2, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs: implicit rules or parallel distributed processing? In J. McClelland, D. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Ward, G. (1997). *Moby pronunciator*. 3449 Martha Ct., Arcata, CA, USA. (Also available at *http://www.dcs.shef.ac.uk/research/ilash/Moby/index.html*)