

**From Rote Learning to System Building: Acquiring Verb
Morphology in Children and Connectionist Nets**

Kim Plunkett

University of Oxford

&

Virginia Marchman

University of Wisconsin, Madison

Abstract

The traditional account of the acquisition of English verb morphology supposes that a dual architecture underlies the transition from early rote learning processes (in which past tense forms of verbs are correctly produced) to the systematic treatment of verbs (in which irregular verbs are prone to error). A connectionist account supposes that this transition can occur in a single mechanism (in the form of a neural network) driven by gradual quantitative changes in the size of the training set to which the network is exposed. In this paper, a series of simulations is reported in which a multi-layered perceptron learns to map verb stems to past tense forms analogous to the mappings found in the English past tense system. By expanding the training set in a gradual, incremental fashion and evaluating network performance on both trained and novel verbs at successive points in learning, it is demonstrated that the network undergoes reorganizations that result in a shift from a mode of rote learning to a systematic treatment of verbs. Furthermore, we show that this reorganizational transition is dependent upon the number of regular and irregular verbs in the training set and is sensitive to the phonological sub-regularities characterizing the irregular verbs. The pattern of errors observed is compared to that of children acquiring the English past tense, as well as children's performance on experimental studies with nonsense verbs. It is concluded that a connectionist approach offers a viable alternative account of the acquisition of English verb morphology, given the current state of empirical evidence relating to processes of acquisition in young children.

Introduction

Classic accounts of the development of English inflectional morphology are typically couched in terms of a three-phased "U-shaped" pattern of acquisition (Ervin & Miller, 1963; Cazden, 1968; Pinker & Prince, 1988). These accounts derive from the following characterization: Children's first productions of English past tense or plural forms are generally correct, regardless of whether those forms are regular (e.g., *walked* or *glasses*) or irregular (e.g., *went* or *sheep*). This correct performance is followed by a period when irregular forms are incorrectly inflected and errors occur, e.g. *goed* or *sheeps*. Finally, the tendency to make such errors decreases, as some forms are identified as exceptions to the predominant regular pattern in the inflectional system.

Abstracting from this somewhat idealized version of the course of development, children's overgeneralization errors serves to support the theoretical assumption that learning a language primarily involves the acquisition of *rule systems*, i.e., explicitly representable generalizations about linguistic regularities which allow the productive generation of forms, including those that are not (or have not yet been) encountered in the input. Further, each of the stages or phases comprising the U-shaped pattern are interpreted as manifesting the application of different *mechanisms* for forming past tense or plural forms, representing different modes or periods within the course of rule acquisition (Brown & Bellugi, 1964; Slobin, 1971; Pinker & Prince, 1988; Pinker, 1991; Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992). While the details of the interpretations vary slightly, most have viewed the first stage in terms of the operation of a *rote* learning mechanism which stores all forms, both regular and irregular, as independent items in a mental lexicon. Here, both irregular and regular forms are correctly produced. In addition, systematic patterns which might characterize the input are not yet encoded, and hence, are not generalized to novel forms encountered by the child.

The second stage marks the child's identification of patterns of regularity, which are explicitly represented as procedures which concatenate stems and inflectional endings, such as "add /-ed/" to form the past tense of a verb or "add /-s/" to form the plural of a noun. It is the overly general application of such rules that results in the production of incorrect forms like

goed and *sheeps*, as well as the tendency to regularize nonsense forms such as *nibbed*. Finally, the third stage involves discovering the exceptions to the rules and isolating the irregular forms as independent entries in a mental lexicon which are outside the purview of the rule mechanism.

The final phase thus supposes the existence of two distinct mechanisms underlying children's ultimate knowledge of the English inflectional morphological system. One mechanism controls the default application of a general rule, responsible for the generativity of the regular paradigm in a given inflectional system. A separate mechanism identifies exceptions and prompts the child to consult its knowledge store when producing and comprehending past tense forms. The most recent incarnation of this dual-mechanism hypothesis explains the phenomenon of overgeneralization using the "blocking principle," i.e., it is when the knowledge store for irregular verbs fails to block the application of the regular rule that overregularizations are assumed to occur (Marcus, et al., 1992). With repeated exposure, the strength of the lexical entry for the irregular verb increases and the tendency for overgeneralization errors correspondingly decreases. The retrieval of irregular lexical items is assumed to involve mechanisms which derive from similar memory-based associative processes that guide the correct early usage of past tense verb forms.

In summary, the onset of the production of erroneous overregularized forms is generally attributed to the transition from a stage in which learning primarily involves expanding the store of individual lexical items, i.e., *rote learning*, to a stage of rule construction and refinement, *system building*. Fleshing out this explanation and providing an adequate interpretation of the phenomenon of overregularization and U-shaped development requires an account of (at least) the following:

- (1) The factor(s) that trigger the transition from rote learning to system building.
- (2) The basis for determining when and how individual lexical items are susceptible (or resistant) to overgeneralization errors.
- (3) The mechanism(s) by which overgeneralization errors are eventually eliminated and appropriate performance is ultimately achieved.

Clearly, the fact that children are able to utilize productively systems of inflectional morphology (not to mention, other aspects of syntax and semantics) is of considerable

theoretical interest. However, as discussed above, most of the evidence suggests that the ability to do so is not evident from the beginning of acquisition, emerging after at least some lexical acquisition has taken place. The factors that trigger this transition in the child have not yet been clearly identified, however, a requisite amount of linguistic experience is often assumed (Karmiloff-Smith, 1986). For example, the onset of usage of the English past tense regular rule is typically thought to depend upon the learning of a sufficient number of suffixed past tense forms. Clearly, without such exposure, it would be difficult for systematicities which define the regular rule to be extracted. However, note that sufficient exposure to non-rule governed irregular forms is also required in order for appropriate blocking to occur.

It has also been proposed that rule-based processes emerge relatively independently of lexical development. For example, the maturation of an inflectional system building device might also determine the timing of the onset of a U-shaped profile of development (Bever, 1982; Pinker, 1991), in particular, one that is associated with the onset of the obligatory marking of tense (Marcus, et al., 1992). Yet note that if maturational factors are found to play a role, it is likely that they interact with input factors to some extent in order to account for observed time lags in the onset of productive behavior in different linguistic domains, e.g., the relatively early acquisition of the English plural system and the typically late acquisition of the past tense system (Brown, 1973; de Villiers & de Villiers, 1985). Other explanations of the time lag in the acquisition of inflectional systems incorporate children's developing conceptual understanding of time and number (e.g., Carey, 1982), as well as the character (e.g., transparency of form-function mappings) of the inflectional system in the language to be acquired (e.g., Slobin, 1985; Johnston & Slobin, 1979).

In interpreting the U-shaped developmental pattern, Plunkett & Marchman (1991) argued that it is important to distinguish between *macro* and *micro* patterns of errors when characterizing children's acquisition of inflectional systems like the English past tense. Macro U-shaped development refers to a rapid and sudden transition into the second phase of system building, resulting in the indiscriminate application of the "add /-ed/" rule to whole classes or categories of verbs. In contrast, a micro U-shaped developmental pattern is characterized by selective suffixation of English irregular verbs, and results in a period of development in which

some irregular verbs are treated as though they belong to the regular paradigm while others are still produced correctly. The basis for selective application of the suffix may be defined with respect to certain representational characteristics of the verb stem (phonological, semantic or otherwise), or may result from the operation of a probabilistic device which determines the likelihood that the suffix will be applied to a given irregular verb.

While a macro view of overgeneralization phenomena has achieved textbook status, there appears to be little empirical evidence that children overgeneralize the /-ed/ suffix indiscriminately, i.e., to all irregular verbs in their current vocabularies (e.g., Maratsos, 1983). Nor is there evidence to suggest the existence of a single well-defined stage of development in which erroneous behavior is observed (see also Derwing & Baker (1986)). Rather, children are likely to overgeneralize the suffix to only some irregular verbs (typically a small number), while at the same time, correct irregular past tense verb forms are also produced. Furthermore, errors may occur across a protracted period with some irregular verbs recovering from erroneous treatment only to be overgeneralized again at a later point in development. Findings undermining the regular rule "imperialism" hypothesis also derive from studies of naturalistic past tense usage (e.g., Marcus, et al, 1992), as well as those using elicitation procedures (Marchman, 1988; Marchman & Plunkett, 1991).

In addition, both children and adults sometimes produce *irregularization* errors (e.g., *flow* → *flew*), in which stem-past tense mappings reflect the sub-regularities characteristic of identity mapping verbs (no change from the stem to the past tense form) and vowel change verbs (Bybee & Slobin, 1982; Marchman, 1988; Marchman & Plunkett, 1991; Plunkett & Marchman 1991). These occur less frequently than the standard "add /-ed/" error, and are more likely to occur in older children and adults. Further, these errors typically (although not always) occur with verb stems that share phonological features of these irregular classes. While most studies report a fairly low frequency of past tense errors in general, errors of both sorts are nevertheless quite pervasive across individuals, as regularized and irregularized forms are observed across extended periods of development in an overwhelming majority of subjects.

In general then, it is fairly well established that the erroneous productions of children learning the English past tense reflect micro-level processes. U-shaped development operates on

a stem by stem basis and is characterized by the selective, gradual, and protracted onset and recovery from erroneous production. Within a dual mechanism approach, it is the operation of the lexical retrieval device (rather than the rule mechanism) that gives U-shaped development its micro character. That is, the onset of rule usage is, by definition, an all-or-none process applicable to any stem, regardless of phonological shape and frequency. Overregularization errors are the result of inappropriate rule application whenever the lexical-based mechanism fails to successfully identify and retrieve an irregular item. The probability of an irregular past tense form being successfully retrieved is dependent upon a variety of factors, including frequency and phonological similarity, that serve to strengthen the status of irregular verbs over the course of learning. In contrast to the absolute nature of the rule-based mechanism, the micro and selective nature of development derives from the associative processes embodied by the lexical retrieval mechanism.

In contrast to the dual architecture assumptions characteristic of rule-based accounts, Rumelhart & McClelland (1986) (henceforth R&M) argued that a single mechanism system in the form of a connectionist network is capable of extracting a range of regularities that characterize the English past tense system and producing patterns of overgeneralizations analogous to the errors observed in children. In their model, the transition from rote learning to system building emerges from the capacity of connectionist networks to simultaneously:

(1) *Memorize* individual patterns and their transformations when the number of pattern types is sufficiently small.

(2) *Generalize* on the basis of regularities observed in the input when the number of patterns (types) is sufficiently large.

R&M initially trained their network on a subset of the vocabulary to which it would eventually be exposed. During the first 10 epochs of training only 10 verbs (8 of which were irregular) were presented to the network. Given the learning and representational resources of their network architecture (a single-layered perceptron), the model succeeded in learning the 10 verbs by rote, i.e., without discovering any regularities among the individual verbs in the training set. After 10 epochs of training, R&M increased the size of the training set by 410 verbs. Consistent with the frequency facts of English, most of these new verbs were regular. Not

surprisingly, this sudden expansion in vocabulary size caused the learning algorithm (a probabilistic version of the Perceptron Convergence Procedure (Rosenblatt, 1962)) to extract the "add /-ed/" regularity and to reorganize the mapping characteristics of the network to reflect the dominant suffixation process. As a result, many irregular verbs displayed a sudden decrement in performance that was eventually overcome with continued training. It is highly likely then that much of the success in modeling the classic U-shaped profile derived from the abrupt manipulation of the number and structure of mapping patterns, and the corresponding transition from item memorization to generalization that is inherent in these networks under such circumstances.

Several critiques (e.g., Pinker & Prince, 1988) noted that the discontinuities introduced into the training regime by R&M do not reflect plausible discontinuities in the input to children. First, there is scant evidence for such an abrupt increase in the total number of verbs to which children are exposed. Second, the evidence from children's productions (Brown, 1973) suggests that the relative proportions of regular and irregular verbs are less skewed than those represented in the R&M training set. For example, Pinker & Prince (1988) note that during early phases of acquisition, regular and irregular verbs are approximately evenly represented in children's production vocabularies. In general, current consensus has targeted the implausibility of the abrupt changes in vocabulary size in the original simulations. Hence the theoretical significance of the U-shaped learning demonstrated by the single mechanism R&M model has been undermined.

More recently, Plunkett & Marchman (1991) demonstrated that several characteristics of micro U-shaped development can emerge in an artificial neural network trained to map verb stems to past tense forms *in the absence of any discontinuities in the training regime*. Here, the network was required to learn an entire set of 500 verbs concurrently. Rather than the result of abrupt changes in the size of the input set, the patterns of errors observed in the Plunkett & Marchman (1991) simulations were shown to derive purely from the competition between the different types of mappings used in the simulations. In particular, overgeneralizations of suffixes to irregular stems resulted from the network's attempt to simultaneously fulfill the constraints of regular and irregular mappings within the confines of a single learning

mechanism. This work also showed that the capacity of these types of networks to learn inflectional verb morphology is highly sensitive to input parameters such as the type and token frequency of stems in the input set, as well as the degree to which the phonological shape of the stem is a predictor of mapping pattern.

Importantly, the errors observed in the Plunkett & Marchman (1991) simulations were predictable in terms of the input factors, frequently bearing an uncanny resemblance to those well-documented in the child language literature. However, the overall size of the vocabulary precluded the network from achieving complete mastery of the vocabulary early in training. Thus, the marked transition from an initial overall high performance to a performance decrement that was achieved in the original R&M model was not observed. Although it was important to demonstrate that competition between mapping types can result in overgeneralization errors in the absence of discontinuity in the input, it is nevertheless unlikely that children attempt to learn an entire lexicon all of a piece. Naturalistic (e.g., Dromi, 1987) as well as parental report measures (Marchman & Bates, in press; Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly, & Hartung, 1992) suggest that verb acquisition in children is a gradual process which follows an *incremental* learning trajectory.

In this paper, we examine the effects of an incremental training regime on networks' ability to learn mappings analogous to those comprising the past tense system of English. (See Elman, 1991 for an application of incremental training to the acquisition of simple and complex syntactic forms). The general training plan is as follows. Early on, the network is exposed to a small number of high frequency stems. In this respect, the simulation resembles the early stage of training in the R&M model. Subsequent to this initial phase, and in contrast to the R&M model, the size of vocabulary is incremented gradually, one lexical item at a time. The selection procedure for adding a particular stem to the training set insures that medium frequency stems are chosen during the middle epochs, while lower frequency stems are added during the later stages of vocabulary growth.

The goal of the incremental training regime is to determine whether a continuous, linear growth in verb vocabulary is adequate to trigger a representational reorganization in the network, i.e., the transition from a purely rote-learning device to a system that can capture the regularities

of verb morphology as well as its exceptions. To this end, network performance is evaluated both with respect to the training set and a set of novel stems. Performance on novel verb stems is particularly important since it reflects the manner in which the network represents the problem domain. The systematic suffixation of novel stems would indicate, for example, that the network has abstracted a generalization beyond a simple memorization of the training set.

We evaluate systematically several parameters associated with incremental training regimes. First, we compare two different incremental training schedules: critical vs. epoch-based expansion. In critical expansion, new verbs are added to the training set only when all previous stem to past tense mappings have been mastered by the network. In epoch-based expansion, new verbs are added to the training set after a given amount of training, irrespective of performance on verbs in the training set. Second, we explore the role of final vocabulary size in determining the degree of generalization to novel stems. There is a natural confound between size of vocabulary and length of training when an incremental procedure is used. Thus, in order to tease these factors apart, we conduct a series of simulations in which vocabulary expansion is halted and training is continued in the absence of vocabulary growth. Third, we evaluate the structure of the training set in determining network performance. It is informative to ascertain the degree to which exceptions to regularities in the training set can block generalization. Performance is evaluated in networks that are trained on early vocabularies that range from consisting exclusively of irregular forms to consisting exclusively of regular forms. Finally, we manipulate the processing resources available to the network by varying the number of hidden units.¹ It is generally assumed within the connectionist literature (Rumelhart, Hinton & Williams, 1986; Hinton, 1989) that increasing the number of hidden units in a network will assist the network in representing additional details of the input and target vectors. On the other hand, decreasing the number of hidden units forces the abstraction of any regularities in the mapping problem. We explore the manner in which these tendencies interact with a mapping

¹ Hidden units receive no direct input from the environment but only from the input units. They provide the network with the capacity to construct internal representations of the input vectors. The similarity of input vector representations at the hidden unit level may differ substantially from the similarity between the input vectors themselves. Hidden units thus introduce a non-linear component into the mapping process.

problem that demands attention to detail (the irregular forms) as well as the opportunity to abstract generalizations (over the regular forms). All of these manipulations permit the assessment of the generality of our findings across different learning conditions (thereby affording a comparison with the circumstances of learning in children) and to evaluate the determinants of network dynamics.

In summary, our primary goal is to determine whether gradual *quantitative* and *structural* changes in the verb vocabulary can lead to *qualitative* shifts in the manner in which a network organizes the mapping relationship between verb stems and their past tense forms. To this end, we hope to demonstrate that an incremental training regime that encompasses both regular and irregular mappings can indeed lead to a representational shift from rote learning to system building within the confines of a single network. The results of these simulations will be of interest to acquisitionists only to the extent that they inform our understanding of the factors that might trigger such a representational shift in young children. Therefore, we provide an assessment of the degree to which the factors that predict the transition to system building in the network are also predictive of the construction of a systematic verb morphology in young children. We will focus on the role of vocabulary size and structure as predictive factors in young children's acquisition of the English past tense and the type of errors observed during the different stages of past tense acquisition.

Method

Overview

All simulations involve training a multi-layered perceptron to map phonologically represented verb stems to their corresponding past tense forms on the RLEARN simulator (Center for Research in Language, UCSD) using a back propagation learning algorithm. Back propagation involves the adjustment of weighted connections and unit biases when a discrepancy is detected between the actual output of the network and the desired output specified in a teacher signal. In multi-layered perceptrons (containing hidden units), error is assigned to non-output units in proportion to the weighted sum of the errors computed on the output layer.

The majority of networks used in this set of simulations contain 18 input units, 30 hidden units and 20 output units (although see *Network Resources* for a description of control simulations where the number of hidden units is systematically varied from 15 to 50). All layers in the network are fully interconnected in a strictly feed-forward fashion.

Training in the simulations follows a pattern update schedule, i.e., a pattern is presented to the net, a signal propagates through the net, the error is calculated, and the weights are adjusted. Pattern update is preferred to batch update (in which error signals are averaged over a range of input patterns before the weights are adjusted) for this problem since children are unlikely to monitor an average error on their output, but are more likely to monitor the error associated with individual pattern tokens. Learning rate and momentum² are held constant throughout the simulation at values of 0.1 and 0.0, respectively. Verb stems are presented

² The learning rate is a constant term in the learning rule for updating connection weights. The size of the learning rate determines the amount a connection changes in response to a given error signal. The momentum parameter is an additional constant term in the learning rule which takes into account changes to the weights made on previous learning trials.

randomly to the network within each epoch³ of training. Performance after selected epochs of training was evaluated using the output analyses procedures described below. All simulations have been replicated with differing random seeds.⁴ However, we have not averaged the results of simulations as this would be inappropriate to the analysis of U-shaped errors.

Vocabulary

A vocabulary of 500 verb stems is constructed from a dictionary of approximately 1000 stems. Each verb in the dictionary consists of a Consonant-Vowel-Consonant (CVC) string, a CCV string or a VCC string. Each string is phonologically well-formed, even though it may not correspond to an actual English word. The dictionary itself is constructed from the 14400 possible CVC, CCV and VCC combinations. However, this number is further reduced by the condition of phonological well-formedness. The final dictionary of 1000 stems is then selected randomly from these remaining well-formed stems. Thus, the base dictionary itself is built from a random sampling of the initial phonological space (subject to well-formedness constraints), contrary to the claims of Prasada & Pinker (1993) that "training and generalization items were drawn from the same small region of the space of possible forms and hence were similar to one another" (p.38).

Each vowel and consonant is represented by a set of phonological contrasts, such as voiced/unvoiced, front/center/back.⁵ Table 1 summarizes the phonological representations for all consonants and vowels used in the simulations.

Insert Table 1 about here

³ An epoch consists of a sweep through the entire training vocabulary. Note that the training vocabulary may contain multiple repetitions of individual verb tokens.

⁴At the start of a simulation all connections in the network are assigned values randomly, typically within the range ± 0.5 . Repeating a simulation but with a different random seed, i.e., a different set of random start weights, permits an evaluation of the degree to which the start state of the network influences training.

⁵ See Plunkett & Marchman (1991) for a more thorough discussion of the phonological representation used here.

Verb stems are assigned to one of four classes. Each class corresponds to a type of transformation analogous to classes of past tense formation in English. The four classes of transformation are:⁶

Arbitrary mappings: There is no apparent relationship between the stem and its past tense form, e.g., "*go* → *went*.."

Identity Mapping: Past tense forms are identical to their corresponding verb stems. Such mappings are contingent upon the verb stem ending in a dental consonant (/t/ or /d/), e.g., "*hit* → *hit*."

Vowel Change: Certain vowels can be changed under the condition that they precede particular consonants. The following four vowel-consonant cluster changes are permitted:

1. /ɪz/ → /eɪz/ "*fɪz* → *feɪz* "
2. /ɪt/ → /eɪt/ "*kit* → *keɪt* "
3. /aɪs/ → /eɪs/ "*laɪs* → *leɪs* "
4. /aɪl/ → /oʊl/ "*rail* → *roʊl* "

Regular mappings: A suffix is appended to the verb stem. The form of the suffix follows the allomorphy of English, and hence, depends upon the final vowel/consonant in the stem:

1. If the stem ends in a dental consonant (/t/ or /d/), then the suffix is /-ɪd/, e.g., "*pat* → *pat-ɪd*."

2. If the stem ends in a voiced consonant or vowel, then the suffix is voiced /d/, e.g., "*dam* → *dam-d*."

3. If the stem ending is unvoiced, then the suffix is unvoiced /t/, e.g., "*pak* → *pak-t*."

The suffixes on the regular past tense forms are represented non-phonologically as three distinct patterns across two output units, i.e., 0 1, 1 0, and 1 1. A fourth pattern (0 0)

⁶ A more fine-grained classification of the past tense of English is provided by Bybee & Slobin (1982) and Pinker & Prince (1988). However, the current four-way distinction serves to capture many of the phenomena of interest.

corresponds to the absence of a suffix, as is the case for stems in the irregular classes (i.e., arbitrary, identity and vowel change).

Stems are assigned randomly from the dictionary to each of the four classes (with no replacement), with the constraint that stems possess the appropriate characteristics of a given class. The resulting 500 verb vocabulary contains 2 stems in the arbitrary class, 458 stems in the regular class, 20 stems in the identity class and 20 stems in the vowel change class. Each of the four vowel-consonant clusters defining the vowel change class contains 5 members. Stem assignments to the arbitrary and regular classes are not contingent upon any particular criteria, and these classes may contain stems which have phonological characteristics of identity mapping or vowel change stems. The total number of stems assigned to each verb class is designed to approximate roughly the verb vocabulary of a child who has already mastered the past tense of English; in particular, regulars greatly outnumber the combined irregular classes, and arbitrary mappings are an order of magnitude lower in number than the other irregular mappings.

Appropriate past tense forms are constructed for each vocabulary item in each of the four classes. In the case of stems in the arbitrary class, a past tense form is chosen that does not share any consonants or vowels with the stem, nor corresponds to the stem or past tense form of any other verb in the training set. The past tense forms for members of the other three classes are constructed according to the criteria listed above.

After 500 verbs have been assigned to the four class types, a subset of 20 verbs is randomly selected from the vocabulary for use in the initial phase of training. In the majority of simulations, the initial training set is comprised of 2 arbitrary stems, 10 regular stems, 4 identity stems and 4 vowel change stems. These initial vocabulary configurations reflect several aspects of what is known about children's early verb vocabularies from naturalistic and parental report measures. For example, data from the MacArthur Communicative Development Inventory: Toddler Form (CDI) (Fenson, Dale, Reznick, Bates, Hartung, Pethick, & Reilly, in press) indicate that of the 20 most frequently reported verbs by parents of children between the ages of 16 and 30 months of age, 10 are regular and 10 are irregular.

The token frequency (i.e., the frequency with which any given stem is likely to be repeated within a single training epoch) during this initial phase of learning is 15 for the arbitrary stems. Regular, identity and vowel change stems have a token frequency of 5.⁷ It has been observed that early verbs learned by children tend to have a high token frequency in the language (Pinker & Prince, 1988). Furthermore, the enhanced token frequency of arbitrary forms reflects previous simulation results (Plunkett & Marchman, 1989) during which such verbs are only mastered by back propagation networks (in the context of a large number of conflicting mappings) when exposure to individual stem-past tense form pairs is frequent.

Criterion versus Epoch-Based Vocabulary Expansion Schedules

The network is trained on the initial vocabularies until all verb stems are mapped correctly to their appropriate past tense forms. Thus, by definition, vocabulary expansion begins at a point in training when performance on the initial set of verbs is perfect. Two general types of expansion schedules were tested. On the first schedule, *criterion expansion*, the vocabulary is expanded one verb stem at a time and trained until that new verb is successfully mapped by the network. In other words, the network must learn each new verb to criterion before having access to other members of the target vocabulary.

On the second training schedule, *epoch expansion*, a new verb is introduced to the vocabulary and trained for a set number of epochs. Another verb is then introduced into the training set. Note that the expansion of the vocabulary occurs *irrespective of the level of performance on the mapping of the previously introduced new verb*. This process is repeated until the vocabulary reaches 500 verbs. Early in training, a new verb is introduced every 5

⁷Pilot simulations indicated that the network would fail to learn all of the regular stems in the initial training set if a token frequency of less than 3 was used. Further increments in the token frequency of regular stems and non-arbitrary irregular stems in the initial training set accelerates their learning relative to the arbitrary stems. However, provided the number of high frequency stems (regular or irregular) is kept small (approximately less than 20) and the number of arbitrary stems does not exceed around 4, the initial training set can be learned to criterion. This robustness in learning of the initial training set permits variability in the relative frequency of the high frequency stems, suggesting some flexibility in the input conditions that support the acquisition of early inflectional verb morphology.

epochs until vocabulary size reaches 100. Thereafter, training is reduced to 1 epoch per new verb. This increased rate of vocabulary growth is intended to model non-linearities in rate of vocabulary growth that are sometimes observed in longitudinal studies of young children (Dromi, 1987). That is, vocabulary expansion proceeds at a relatively slow pace early on, while later growth is more rapid.

The order in which new verbs enter the vocabulary is determined by a weighted random selection process based on an 80% likelihood that the new verb is taken from the regular class and a 20% likelihood that the verb is taken from the identity or vowel change classes. (Recall that the two arbitrary verbs are members of the initial vocabulary.) Each new verb entered into the training set after the initial set of 20 is assigned a token frequency of 3, until the vocabulary size reaches a total of 100 verbs. Thereafter, verbs that are introduced (predominantly regulars) are trained using a token frequency of 1. This frequency profile was again chosen to accommodate the data set to the observation that children are more likely to hear, and thus have a greater opportunity to learn, verbs with a high token frequency.

A summary of the changing structure of the vocabulary by verb class is provided in Table 2, and Figure 1 plots the relative proportion of regular and irregular verb tokens as vocabulary size is expanded. Notice that Figure 1 indicates a switch in early training from a predominance of irregular verbs (approximately 60%) to a predominance of regular verbs. This switch reflects a changing proportion of regular and irregular verbs that is observed in children's early vocabularies as reported by Marchman and Bates (in press) and contradicts the claim by Pinker and Prince (1988) that the proportion of regular and irregular verbs in children's early vocabularies are approximately equal throughout development.

Insert Table 2 about here

Insert Figure 1 about here

Initial Vocabulary Structure

While child language data for English suggest that early verb vocabularies are likely to undergo a shift from a predominance of irregular to regular verbs, it is also known that considerable individual variation exists in early vocabulary composition (e.g., Bates, Bretherton, & Snyder, 1988). Hence, we conducted several series of simulations that vary in terms of their initial proportion of regular to irregular verbs. Nine early vocabulary configurations are tested using an epoch-based expansion schedule. Each configuration differs in terms of the number of regular forms to which the network is exposed during the earliest phase of training. For example, the number of regular and irregular forms comprising the initial training set of 20 verbs in each of the nine conditions varied from 0 regulars-20 irregulars to 18 regulars-2 irregulars. In four of the conditions, the proportion of regulars to irregulars remains approximately constant until the 40 verb mark (0%, 25%, 50%, 75%). In the remaining 5 conditions, the proportion of regulars either decreased or increased as vocabulary expanded from 20 to 40 verbs. Thereafter, for all conditions, vocabulary expansion proceeded according to the weighted random selection procedure described above, and token frequency parameters were assigned in the standard fashion. (The exact numbers of items in each of the vocabularies are presented below accompanying the discussion of the results.)

These control simulations permit an evaluation of (1) the degree to which the presence of regular verb forms early in training is a necessary condition for the subsequent onset of generalizations, and conversely, (2) whether an overabundance of irregular forms early in training can *block* the development of system building in the network and hence the generalization to new regular forms.

Network Resources

Given the relationship between generalization abilities and number of hidden units in these networks, the role of representational resources was explored by manipulating the number of hidden units used to configure the architecture of the network. Two measures of performance were of interest here: (1) the ability to learn verbs in the training set, and (2) the networks' ability to generalize to novel verb forms. A total of 9 conditions were evaluated,

ranging from using 10 hidden units to 50 hidden units, in increments of 5 units. For each condition, the epoch-based expansion schedule was used, i.e., after training on an initial set of 20 verbs (10 regular and 10 irregular), a new verb was added to the training set every 5 epochs. The same weighted random selection procedures and token frequency parameters as described above were used. Hence, the only factor varied across condition was the number of hidden units.

Novel Verbs

A set of 100 legal stems which were not included in the training set were selected from the dictionary for testing the generalization properties of the network. Of these, 10 end in a dental final consonant (/t/ or /d/): Identity Mapping; 10 stems possess the characteristics of each of the 4 clusters defining the vowel change class (a total of 40 stems): Vowel Change; and 50 stems that did not possess any of the previously mentioned characteristics: Indeterminates. It is worth emphasising that the Indeterminate novel stems do not form a well-partitioned group in phonological space. They have no more phonological features in common with each other than they do with the Identity Mapping or Vowel Change novel stems. These sub-classes of novel stems permit an evaluation of the manner in which the network has tuned its response characteristics to the presence (or absence) of specific phonological features in the stems making up the training set. The network's performance on the novel verbs was evaluated at regular intervals during training using the output analyses procedures described below.

Output Analysis

The weight matrices were saved at regular intervals; first, when the net had just mastered the initial 20 verbs and then each time a new verb was introduced but before any training on the new verb had occurred. These weight matrices provide snapshots at various points in training that permit evaluations of the accuracy of the network in producing the correct past tense form for each unique stem at different points in development. For every given stem, the output of the network was evaluated in terms of the "closest fit" (in Euclidean space) to the set of phonemes that map the output space, defined by the teacher signal to the network (see Table 1). Error analyses provided an overall hit rate (i.e., % correct), as well as the proportion of stems in each

class that were regularized (i.e., add a suffix), as well as irregularized (i.e., incorrectly mapped as identity stems, vowel change stems, blends, etc.). The following error coding categories are used:

SUF: The stem is regularized. For regular stems, this indicates that an inappropriate but otherwise legal suffix is affixed.

ID: The stem and past tense have the same form.

VC: The stem undergoes a vowel change. For vowel change stems, this indicates that an inappropriate vowel change occurs.

BLD: The stem is blended, i.e., it undergoes both vowel suppletion and suffixation.

UNC: Unclassifiable responses (typically incorrect mapping of consonants).

Novel verb stems were also tested on each of the saved weight matrices at each testing point. Using a similar output analysis procedure, the three different categories of novel verbs (indeterminate, identity, vowel change) were analyzed separately to determine their output tendencies, i.e., whether they tend to be regularized, irregularized or handled in some other fashion by the net as a function of amount of training and vocabulary size.

Results

Criterial Expansion

We begin our results presentation by outlining the overall ability of the networks to learn when vocabulary expansion is contingent upon previous performance (criterial learning). Recall that in this condition, vocabulary size was increased one verb at a time and training continued on each new verb (as well as the initial set) until that new verb was successfully mapped by the network. The results indicated that training on the initial set of 20 verbs required approximately 15 to 40 epochs to reach criterion, depending on the initial configuration of random weights. Further, successful training on subsequently added stem-past tense mappings consistently failed when vocabulary size reached approximately 27 verbs. In other words, criterial expansion appeared to fall considerably short in its ability to allow the network to structure its resources in such a way so as to master the entire set of 500 stems in the target vocabulary. In order to verify that subsequent learning was impossible, training was continued for a considerable number of epochs in each case. Analyses of the mean square error on the output units clearly indicated that the error reaches asymptote at a non-zero level in all networks at around this vocabulary size.

The inability of networks trained using the criterial expansion procedure to learn a large number of verb stem/past tense mappings reflects the propensity of networks of this type to be caught in "local minima". Learning in networks of this type can be understood as the process of traversing a hilly multi-dimensional landscape where the regions of any part of the landscape are defined by the values of the weight matrix and the height of the landscape is just the error that results from a given configuration of the weight matrix. For example, in a network with just two weights we can define an error surface in three dimensions where the error is plotted on the vertical axis and the coordinates in the horizontal plane are just the corresponding values of the two weights. A local minimum is a point on the error surface where the gradient is zero but which does not correspond to the global minimum of the error function. Since the learning algorithm used in the network is sensitive to the slope of the error surface, certain configurations of the weight matrix can result in the network becoming entrenched in a given

state where no further learning can occur. Alternatively, the weights feeding into a given unit may become very large (either positive or negative) as a result of repeated presentations of the same input. Other inputs subsequently presented to the network which use the same weight lines must fight an uphill battle to overcome the mapping characteristics of the previously presented input. Because early inputs to the network continue to be presented with a relatively high frequency, new inputs will have difficulty overcoming the early bias of the network. It is clear, therefore, that the mapping characteristics of the initial training set can have important implications for subsequent learning when a criterial incremental learning schedule is used in network architectures of this type. In order for networks to avoid entrenchment in inappropriate areas of weight space, training must ensure that a variety of weight changes occur. If the network is repeatedly trained on a limited and fixed number of patterns, where a series of similar weight changes occur, further training may fail to promote necessary reorganizations or may even enhance the network's entrenchment in a particular region in weight space. This training schedule was therefore abandoned as a method of vocabulary expansion that is appropriate to the current task.

The following sections report on results from simulations in which verbs are added to the vocabulary *irrespective* of the level of performance of the network on the previously added verb.

Epoch Expansion

Overall Performance

Figures 2a and 2b summarize performance (% correct) on verbs in the irregular (arbitrary, identity and vowel change combined) and regular classes, respectively, as a function of vocabulary size.

Insert Figure 2 about here

Recall that before vocabulary expansion is allowed to begin, the network is trained to 100% accuracy on the initial set of 20 verbs. There are several things to note in Figures 2a and

2b. First, overall performance (as measured by hit rate for all current trained forms) deteriorates substantially at several points in training. Nevertheless, the network eventually recovers from these setbacks. For example, by the time vocabulary reaches 500 verbs, irregular verbs have achieved 100% correct output, and between 95% and 100% of the regular verbs were produced correctly.⁸ In general then, the epoch-based incremental training schedules were considerably more successful at allowing the networks to master the entire vocabulary than criterial based learning schedules. It should be stressed that the decrements in performance plotted in Figure 2 are primarily the result of network inaccuracies in mapping *new verbs* that are entered into the training set. The decrements *do not* necessarily indicate the unlearning of verbs which had already been mastered by the system. Many of the verbs in the training set may continue to be mapped appropriately (cf., the criterial expansion schedule above), while others may indeed be "unlearned," demonstrating a sort of U-shaped development. Analyses of the patterns of U-shaped learning in these simulations are discussed below.

A closer look at Figure 2 reveals two major periods of decrement. For both regular and irregular verbs, overall performance drops fairly early in learning, almost immediately after vocabulary size has begun to increase. Yet, recovery comes quickly, first manifesting itself for the regular verbs when total vocabulary size reaches approximately 44 verbs, and for the irregulars at approximately 31 verbs. Given that verbs are introduced into the vocabulary at a constant rate both before and after these periods, these data suggest that beyond a vocabulary size of around 50 items, new verbs appear to be learned faster than verbs introduced during the early stages of expansion. Thus, there is preliminary evidence to suggest that the number of verbs in the current vocabulary may indeed be an important factor in determining the network's ability to learn new lexical items, and presumably, to generalize to novel forms. We will continue to evaluate this hypothesis in subsequent sections.

Figure 2a also indicates that irregular verbs undergo a substantial decrement in overall performance during the middle period of training, i.e., when vocabulary size reaches approximately 125 verbs. During this period (125 to 210 verbs), the number of irregular verbs

⁸Absolute final level of performance on regular verbs varied within this range as a result of variations in the initial weight matrix for different simulations.

in the vocabularies has increased from approximately 23 items to approximately 37 items (including an extra 5 IDS and 9 VCS). An analogous, but much less drastic, decrement in overall performance is observed for the regular verbs (Figure 2b) around the same period (when total vocabulary size ranges from approximately 90 to 260 verbs). In interpreting these data, it should be recalled (again) that these decrements in performance do not necessarily indicate the U-shaped "unlearning" of individual past tense mappings. These patterns of performance are also a reflection of the inability of the system to learn new items that are entered into the vocabulary. Indeed, this interpretation relates directly to the changes in the input frequency characteristics of verbs introduced after the 100 vocabulary mark. Recall that all verbs (both regular and irregular) that are introduced after the 100 vocabulary mark are trained with a token frequency of 1. Previous work with networks of this type (Plunkett & Marchman, 1991) has shown that mappings with a low token frequency (in particular, irregular arbitrary and vowel change verbs) are difficult for a network to master in the context of a large number of conflicting mappings.

In general then, mastery of this set of mappings does not follow a straight-forward learning function in the context of gradual and incremental increases in vocabulary size. In subsequent sections, analyses seek to isolate the degree to which decrements in overall correct performance reflect the inability of the system learn to new items entering the vocabulary, in contrast to the "unlearning" of forms that were previously successfully mapped. More specifically, we attempt to target whether changes in overall performance reflect *qualitative* changes in network organization deriving from these incremental changes in vocabulary size.

Errors as a function of vocabulary size

Network output for every stem in the current vocabulary is determined using a closest-fit algorithm and all incorrectly generated mappings are categorized by verb class. A verb stem is incorrectly mapped when the closest fit of the output (in Euclidean space) for each of the phonemes does not match that specified in the teacher signal. For these analyses, errors include both inappropriate output on verbs that had yet to be correctly mapped by the network, as well as incorrect mappings for verbs that had previously been successfully mastered by the network. Table 3 summarizes the frequency and timing of errors as a function of expanding verb

vocabulary size (represented by successive rows in the table), as well as hit rates (% correct) for each verb class. An analysis of the arbitrary class is not presented since these items perform at optimal level throughout the expansion schedules. We review evidence below (from Marcus et al. 1992) that indicates, contrary to popular belief, that errors of the kind 'go-ed' and 'went-ed' (overregularizations to the arbitrary class) are very infrequent in children's spontaneous speech. Given the constraints on learning inherent in these networks, arbitrary stems would become more susceptible to error if their token frequency had been lower, e.g. 5 instead of 15 (Plunkett & Marchman, 1991). The robustness of the arbitrary class in the current set of simulations reflects the functional modularity that can be achieved within the confines of a single mechanism that must learn to perform multiple types of mapping via the manipulation of frequency characteristics of the input. Error scores indicate percentage of total errors for items in a given class.

Insert Table 3 about here

Several comments should be made concerning the data presented in Table 3. First, note that the overall level of errors is low, as the *average* hit rate across the training period is 95.6% for regulars and 97.6% for irregulars. Second, when errors did occur, they were likely to be circumscribed to a limited range of error types for all classes. The error categories of **SUF**, **ID**, and **VC** account for the overwhelming majority of incorrect responses. Residual errors derive primarily from the incorrect mapping of consonants on regular verbs. These unclassifiable responses (**UNC**) were more common early in training, although a few did occur as late as a vocabulary size of 490 verbs. Recall that the network is forced to make a response on every trial, with the output determined by the closest fit in Euclidean space to legal phonemes. It is also possible to restrict the output of the network, and hence eliminate many of the unclassifiable responses, if the closest fit metric is supplemented by a proximity criterion (i.e., closest phoneme and within a specified distance). In this case, we would be evaluating only those responses for which there was some degree of certainty, analogous to a child using a particular past tense verb form only when he/she is relatively sure of how to pronounce all components of

that form. Interestingly, however, current evidence from the child language literature suggests that eliminating unclassifiable forms in this fashion may not be entirely valid. For example, Plunkett (1993) reports the use of non-standard forms by children which involve the substitution of inappropriate vowels and consonants in target lexical forms. Admittedly, these non-standard usages are most prevalent in that period of language development prior to the vocabulary spurt. However, incorrect usage of phonemes in words did not disappear entirely from children's later productions. (See Cottrell & Plunkett (1991) for further discussion.)

It should also be noted that although Identity Mapping errors can be unambiguously identified in the network's performance, this is not the case for young children. In the network, every unchanged output is by definition an Identity map, and it is an Identity Mapping error when the form in fact required a change. In contrast, in children, no change to the stem may result from a process of Identity Mapping *or* a failure to include tense marking in an obligatory past tense context. It is difficult to ascertain, in English-speaking adults or children, which of these two accounts applies to any individual no change error. We will follow other researchers (Brown, 1973, Bybee & Slobin, 1982, Marchman, 1988) in assuming that when children mark for past tense in over 90% of obligatory contexts, then responses which are identical in form to the stem can indeed be considered to be Identity Mapping errors. Our model takes no account of the role of children's conceptual development regarding tense or pastness in explaining their past tense errors.

Each class of verbs was susceptible to different types of errors. Table 3 indicates that items belonging to the identity mapping class were likely to be produced correctly throughout the training period (M=98.4% correct), yet some errors did occur (maximum error rate = 16% of responses). Of the errors that were produced on stems in this class, the responses were primarily erroneous suffixations (M= 14%) and to a much lesser extent, vowel changes (M= 3%). Unclassifiable errors were observed on only one occasion. In addition, it is interesting to note that all of the identity stems that underwent an erroneous vowel change possessed the requisite vowel/consonant stem final combination, and identity stems *never* underwent blending (i.e., a simultaneous vowel change and suffixation).

Vowel change stems were also mapped successfully across a large part of the training period (M=96.7% correct), yet a range of errors occurred during a fairly circumscribed portion of learning (vocabulary sizes between 130 and 210 items). If an error was produced, it was most likely to be an identity mapping (M= 5%), inappropriate vowel change (M= 3.7%), or blend (M= 6.8%). Suffixation errors were observed on a few occasions, although these comprised less than 1% of the responses to vowel change stems. Unclassifiable responses were rare, but not completely absent.

While regular verbs were incorrectly mapped only an average of 4.4% of the time (range = 0 to 20%), when an error did occur, it was likely to result from identity mapping (M= 18%), an inappropriate suffix (M= 18%), or a blend (M= 10%). Note that regular stems were susceptible to identity mapping and inappropriate suffix errors throughout the training period, however, blends were more likely to occur during the last part of training. Further analyses indicated that while most of the regular stems that were identity mapped ended in a dental consonant (M= 85%), a small proportion of non-dental final regular stems were also identity mapped (M= 15%).

In summary, a gradual, epoch-based schedule of increases in vocabulary size allowed the network to master the entire set of mappings, yet distinct periods of erroneous performance were still observed. Interestingly, certain error types were likely to occur with verbs that possessed particular characteristics, suggesting that errors were partially conditioned by the phonological shape of the stem. This was true for stems which underwent an inappropriate identity mapping (i.e., ended in a dental consonant), as well as those undergoing vowel suppletion. Further, unclassifiable responses were more likely to be observed early in training, while error types that resulted from the merging of response patterns (i.e., blends) were more likely to occur later. In the next section, we focus on analyses of the relationship between these errors types and previous performance – i.e., the nature and timing of "unlearning" in these networks.

Regularizations, Irregularizations, and U-Shaped Development

We continue our analysis of network performance by outlining the nature and timing of errors produced by the network in relation to what is known about error production in children. In discussions of past tense acquisition in children, it is customary to focus on a particular type of error, *regularizations*, i.e., errors in which an irregular stem is treated as though it were regular in its past tense form. Following Marcus et al. (1992), these include errors on irregular verbs which have previously been correctly produced by the child, as well as those that have not yet been produced correctly in their past tense form. Using the same criteria as Marcus, et al (1992, p. 29), we derive the rate of regularization using the following formula:⁹

$$\frac{\# \text{ overregularization tokens}}{\# \text{ overregularization tokens} + \text{correct irregular past tokens}}$$

Figure 3b plots the frequency of overregularization of irregular verbs in the current set of simulations as a function of vocabulary size. For purposes of comparison, Figure 3a presents data on the frequency of overregularization errors for one child, Adam, as reported in Marcus et al (1992) (Figure 3, p. 38).

Insert Figures 3a and 3b about here

The important characteristics of the pattern of overregularization errors evident in both Adam (Figure 3a) and the network (Figure 3b) include:

- (1) A generally high level of performance on the past tense forms of irregular verbs across the period.
- (2) An initial period of error-free performance on irregular past tense forms.
- (3) A prolonged period where a small minority of irregular verbs are overregularized (i.e., suffixed), including both irregular verbs that have been previously mapped

⁹ Note that in the simulation data, this measure reflects the number of irregular verb types (not tokens) that are overregularized. Because the network is asked to produce the past tense form of each verb in the training set exactly once per testing, we have no way of assessing the frequency with which a given individual stem will be produced erroneously.

correctly by the network (i.e., that have undergone U-shaped unlearning), as well as irregular verbs that have been introduced recently to the training set.

This general pattern is characteristic of data reported by Marcus, et al. (1992) for other children in the Brown corpus (i.e., Eve and Sarah, their figures 4-5, pg. 38-39). However, this profile is different from that reported for Abe (see their figure 6, pg. 39), who does not show an initial period of correct performance on irregular past tense forms. Hence, while the classic pattern of overregularizations following initial correct performance may be exemplified by many children for whom longitudinal data are available, it may not be characteristic of the learning pattern of all children and may not reflect the operation of a separable, rote learning device.

In order to investigate whether the general pattern reflects the U-shaped unlearning of individual stems, we isolated those stems that were correctly produced by the network, then incorrectly mapped at some subsequent point in training, and then finally, correctly output once again. A total of 15% of the Identity mapping and 30% of the Vowel Change verbs were classified as undergoing U-shaped acquisition according to these criteria. None of the arbitrary stems underwent U-shaped learning, as both forms were produced correctly across the period, but such errors are apparently very infrequent in the spontaneous productions of children (Marcus et al., 1992). For the dental final stems, U-shaped errors were most likely to result in the overregularization of the stem (66%); however, one of the dental final stems was *irregularized*, in particular, treated as if it were a member of the vowel change class. For vowel change stems, many of the U-shaped errors resulted from a blending of a vowel change and a suffix (50%, e.g., *lls* → *lAst*). Only one pure regularization error occurred on a vowel change stem throughout training. Another source of U-shaped errors for this class of stems was identity mapping (50%). This pattern of U-shaped errors for Vowel Change verbs does seem to diverge from the pattern of pure regularization errors on Vowel Change verbs in children; that is, examples of overgeneralizations often involve vowel change stems, e.g., *comed*, *seed*, *blowed*, *breaked*, *winned*. While the source of this discrepancy is as yet unknown, it is important to point out that information regarding the relative frequencies of error types as a function of a verb class in children is very sparse.

U-shaped errors were also analyzed with respect to when they occurred, i.e., at what point in training after the verb was correctly mapped was the first incorrect output observed. This analysis indicated that irregular verbs entering the lexicon both early and late in training were susceptible to U-shaped acquisition. The majority of U-shape onsets for irregular verbs (6 of 9, or 66%) occurred during the first half of training (i.e., vocabularies less than 250 verbs). However, only one irregular verb in the initial set of 20 was U-shaped. Given that all of the irregular verbs have been entered into the vocabulary by the 250 verb mark, these data indicate that verbs entering the vocabulary during the middle portion of training were most susceptible, and were overregularized fairly soon after they were first mastered. Indeed, the last U-shape on an irregular verb occurred at a vocabulary size of 375 verbs. The error data (see Table 3) also indicated that irregular stems were likely to be mapped correctly during the second half of training.

Regular stems were also susceptible to U-shaped learning, i.e., produced correctly and then "unlearned" at a later point in training. However, as in children, these were likely to occur with only a small subset of the total number of regular verbs (17%). These errors were most likely to result from identity mapping (M=35%), blending (M=24%) or the addition of an inappropriate suffix (M=22.5%). All inappropriate suffixes were, of course, legal suffixes (closest suffix in Euclidean space). To our knowledge, the extent to which children in this age range add the wrong allomorph of /ed/ to their past tense verb productions is not known. Yet, it has been reported that adults make similar non-standard voicing assimilation pronunciations in on-line tests of past tense production, e.g., /spel-t/, /spil-t/ (Marchman & Plunkett 1991), and children often produce inappropriate suffixes (of the /t/ and /d/ types) in spelling tasks (Peter Bryant, personal communication). U-shape errors on regular verbs were observed throughout the training period. However, like the irregular verbs, approximately two-thirds occurred when vocabulary size was less than 250 verbs. As the vocabulary increased beyond this point, U-shaped errors were more likely to result from the incorrect mapping of consonants (18%).

In summary, the erroneous performance in these networks is the result of errors on new forms entering the vocabulary, as well as the "unlearning" of previously learned verbs. The behavior of the network resembles what is known about children in several respects. In both

networks and children, erroneous output is relatively infrequent (compared to correct performance), although both regularization and irregularization errors are observed across a large portion of the training period. The U-shaped overregularization of irregular forms is most likely to occur during the early and middle portions of training, while blends (e.g., *ated*) and other irregularization errors are likely to persist later in acquisition.

Responses to Novel Stems

The preceding analyses evaluated changes in the network's ability to produce the appropriate past tense forms of verbs that were members of the training set. Here, we investigate network performance when it is required to produce the past tense forms of stems that it has never seen, i.e., novel verbs. As with children (e.g., Berko, 1958), the ability to generate reasonable past tense forms of novel stems provides a measure of the extent to which the network has abstracted useful information from the training set. Further, the way in which these tendencies *change* over the course of acquisition can be seen to reflect concomitant representational changes, i.e., reorganizations, within the network.

In Figure 4, we graph the output of the network when presented with three classes of novel forms. Figures 4a, d, and g plot the tendency of the network to treat Indeterminate novel stems as if they were members of the regular, no change, or vowel changes classes, respectively. Figures 4b, e, and h plot the same tendencies for novel stems which possess a dental consonant in stem final position (Identities). Finally, Figures 4c, f, and i present responses for novel stems possessing the VC clusters characteristic of vowel change verbs (Vowel Changes). The extent to which the network produces systematic responses to verbs with each of these characteristics indicates its sensitivity to the presence (or absence) of phonological characteristics which are predictive of class membership in the training set.

Insert Figures 4a to 4i about here

Note the high level of regularization in response to indeterminate novel stems (92%) by the end of training as presented in Figure 4a. While suffixation is rare at the onset of training, it

rapidly increases yielding an average suffixation rate of 71% across the period. In contrast, the average rate of producing identity mapping forms in response to indeterminate stems is very low (3%) – see Figure 4d. In a similar fashion, virtually none of the indeterminate stems undergo vowel change mappings except for a temporary blip early in training – see Figure 4g.

Identity novel stems were subjected to the entire range of generalization tendencies, resulting in an average of 18% suffixation, 27% identity mapping, and 26% vowel change responses. Interestingly, however, these responses were not equally probable across the course of training. For example, the tendency to add a suffix to an identity stem was quite strong early in training (i.e., when vocabularies are smaller than say, 200 verbs) but then decreased, whereas the frequency of identity mapping and vowel change responses started off at modest levels and tended to increase over the same period.

Like Identity stems, Vowel Change novel stems were subjected to the entire range of response types, although the most predominant pattern was to treat these novel stems in accordance with their phonological shape, i.e., as if they were members of the vowel change class (M=38%, see Figure 4i). Early in training, this tendency increased slightly when vocabularies reached between 100 and 200 verbs, but remained fairly steady across the rest of training. VC stems were also identity mapped to a certain extent (11%). Unlike the Identities, there was not an early overwhelming preference for vowel change stems to be suffixed. Instead, this tendency remained fairly constant across the period at approximately 28%.

These patterns suggest that the network is highly sensitive to the phonological properties of stems when generating the past tense forms of novel verbs. Novel stems which possess phonological properties characteristic of the sub-regularities of the irregular training stems are more likely to undergo a past tense mapping associated with the irregular class (i.e., irregularization) than any other type of mapping. Thus, Identity novel stems are more likely to undergo identity mapping than suffixation or vowel change; Vowel change novel stems are more likely to undergo vowel change than identity mapping or suffixation. Yet, note that these mapping tendencies are not absolute. Identity and Vowel Change novel stems also undergo past tense mappings characteristic of the other irregular classes, as well as the predominant regular class. This is partly because class membership is not absolute. In the training set, the regular

class possessed members that have characteristics associated with the irregular classes, i.e., dental final stems or VC clusters. Similarly, irregular verbs in the training set overlapped in their phonological properties. For example, one of the phonological features of a VC cluster was /it/ → /ɛt/ as in "kit → kɛt". In this case, a VC cluster shares a property associated with the identity mapping class, i.e., a dental final consonant.

Nevertheless, the propensity to regularize indeterminate novel stems regardless of their phonological shape suggests that the trained network has effectively defined a *default mapping strategy*. Recall that the indeterminate stems used to test the network do not form a well-partitioned group in phonological space. They are no more similar to each other than they are to the other classes of novel stems. The net result is that novel stems which do not possess the phonological features characteristic of other mapping types will be treated in a similar fashion. Within the context of this artificial language that is structured to look very much like English, these stems are suffixed. It should be stressed, however, that the tendency to regularize indeterminate stems is not absolute and categorical. For example, an average of 3% of the indeterminate novel stems undergo identity mapping, even though these stems do not possess the requisite phonological features. This response tendency is another consequence of the fact that phonological shape is characteristic, but does not necessarily define each of the original irregular classes. Features of the various classes overlap, yielding some regular stems in the training set that end in a dental consonant and a few that possess the CV cluster typical of Vowel Change verbs. Nevertheless, the regularization process is very strong, especially during the early and middle periods of development – so strong, in fact, that more than one-fifth of the identity and vowel change novel stems undergo suffixation.

In summary, these data suggest that the network has learned to regularize novel stems in much the same fashion as children regularize nouns and verbs in a "wug test" (Berko, 1958). In the absence of phonological cues, the network overwhelmingly chose to apply the regular pattern, and did so to a certain extent even when phonological cues suggested an alternative response. However, in many cases, phonological features of the novel stem *can* block application of the suffixation process, resulting in the *irregularization* of novel stems (e.g., *flow*

--> *flew*). Similar findings have also been reported for children (Bybee & Slobin, 1982; Marchman, 1988; Marchman & Plunkett, 1991).

Generalization as a function of vocabulary size

Interestingly, the tendency to regularize indeterminate novel stems alters dramatically between early and late training. Early in training, i.e., immediately after the network has mastered the initial vocabulary of 20 verbs, indeterminate novel stems were treated in an unsystematic fashion. Network output was unclassifiable in terms of the mapping categories used in Table 3. During this early period of training, only one of the novel indeterminate stems was treated in a systematic fashion. It was regularized. The tendency to systematically treat identity and vowel change novel stems was greater, including 3 regularizations, 3 identity mappings and 11 vowel changes. These data suggest that although the network has not yet extracted a mapping from the limited training set that it will apply in a default fashion, it has already begun to develop a sensitivity to the phonological characteristics of the irregular verbs.

It is difficult to evaluate these network findings against empirical data from children at an equivalent level of language development. Any such evaluation would involve testing early two-year olds on novel verb forms which are systematically manipulated with respect to phonological form. These experiments have not been performed. However, where such experiments have been carried out on slightly older children (Bybee & Slobin 1982, Marchman 1988) there is clear evidence of sensitivity to the sub-regularities that characterize irregular forms, e.g., dental-final forms are *less* likely to be overregularized than the other sub-classes of irregulars, especially early in development. Consistent with this view, one might postulate that the discovery of sub-regularities in the initial verbs learned by children would help consolidate the learning of these verbs and contribute to maintaining the early period of error free performance.

As vocabulary size expands (and training continues), the tendency for the network to add a suffix to indeterminate novel stems increased substantially (from 2% to 92%). The most rapid increase occurs during the period of training when vocabulary size increases from 30 to

140 verbs. A major proportion (76%) of indeterminate novel stems are regularized by the 140 vocabulary mark. Note that this point in training corresponds to a period in which performance on *trained* irregular verbs also deteriorates (see Figure 2). Thereafter, the rate of increase in regularization of novel indeterminate stems is seen to decelerate.

The relatively sudden onset of the systematic treatment of novel stems suggests that abrupt reorganizational processes are occurring in the weight matrix of the network. However, it is unclear whether these changes are the result of prolonged training or the network's exposure to an increasing number of different stems as vocabulary expansion continues. In order to tease apart these two factors, we trained the network on different levels of final vocabulary size. In these test simulations, training proceeded in precisely the same fashion as in previous simulations except that vocabulary expansion was halted at 6 different levels – 30, 40, 50, 60, 70 or 80 verbs. After vocabulary expansion was stopped, training continued until the 500 epoch mark. Figure 5 plots the frequency with which indeterminate novel stems were mapped as regular past tense forms in the different simulations.

Insert Figure 5 about here

These generalization curves indicate that final vocabulary size, rather than amount of training (number of epochs), is the best predictor of final level of generalization. That is, the ability of the network to generalize the suffix to indeterminate novel stems remained at a low rate when final vocabulary sizes are small, regardless of the amount of training. Interestingly, a substantial increase in generalization tendencies is observed only when vocabulary sizes increased beyond the 30 verb mark. Hence, the relationship between changes in final vocabulary size and final level of generalization is non-linear.

These control simulations point to a strong relationship between size of vocabulary and the onset of the tendency to apply the default mapping to novel forms. Continued training on a data set that is comprised of an insufficient number of exemplars does not result in the abstraction of the generalization to the degree that is observed in children who are acquiring systems like the English past tense. This relationship between vocabulary size and onset of the

'default' mapping predicts that there will be a certain amount of continuity between rate of development in the lexicon and inflectional morphological systems. There is some evidence to suggest that this relationship also obtains in children (Marchman & Bates, in press), although Marcus et al (1992) argue that vocabulary size does not predict the onset of overregularization errors. We will return to these findings in the discussion section and attempt to show how these discrepant reports can be resolved within the framework of a connectionist account of the acquisition of inflectional morphology.

Vocabulary Structure

The results reported above indicate that the tendency of the network to regularize novel stems is primarily a function of the total number of vocabulary items upon which the network has been trained. However, in all the simulations presented thus far, initial vocabularies were configured using an equal proportion of regular and irregular verbs, i.e., 10 regulars and 10 irregulars. In this section, we briefly summarize the results of a series of simulations in which we systematically vary the structure of this initial vocabulary configuration. Table 4 provides an overview of the nine vocabulary configurations that are evaluated, the relative proportions of regular to irregular items when each vocabulary reaches a total of 40 verbs, as well as the proportion of indeterminate novel stems which are suffixed at the 40 verb mark. These series of simulations seek to determine whether it is the particular initial vocabulary configuration (i.e., 50% regular, 50% irregular) that is responsible for the relationship between lexical and morphological acquisition observed above. More specifically, we assess whether the onset of generalization ability to novel stems is best predicted by: (1) total size of vocabulary (regardless of its composition), (2) total number of regulars (regardless of the number of irregular forms), or (3) the relative proportion of regulars in the training set (regardless of total size).

Insert Table 4 about here

Using the frequency of regularizations to indeterminate novel stems as the variable to be predicted, multiple stepwise regressions were performed within and across vocabulary structure

conditions. Taking all nine initial conditions together, analyses indicated that total number of regulars and percentage of regulars were each significant predictors of generalization behavior, contributing 78% and 58% of the variance, respectively, when entered into the regression equation first. Total vocabulary size also correlated significantly with proportion of indeterminate stems regularized ($r = 0.63$). However, this correlation was not as strong as the correlation between regularization and total number of regular verbs or proportion of regular verbs ($r = 0.87$, $r = 0.76$ respectively).

Estimates of the unique contribution of each variable were also obtained by entering each into the equation last. Looking first at total number of regulars, this procedure indicated that this variable made a unique contribution of 23% of the variance. In contrast, *proportion* of regular forms accounted for only 5% of the overall unique variance in the regularization of novel forms. This difference in performance resulted from the fact that generalization was consistently poor (<5%) whenever percent regulars dropped below 50% of the total vocabulary. This situation occurs only in conditions 1→4, and as long as total vocabulary size has not expanded beyond 85 verbs. Thus, while the absolute number of regulars in the vocabulary is a better predictor of generalization than percent regulars in the vocabulary, this is due primarily to the fact that generalization is virtually absent when regulars contribute less than 50% of the items overall.

In summary, the level of generalization observed in these simulations is closely related to the total number of regular verbs in the vocabulary, provided the proportion of regulars exceeds approximately the 50% level. In those vocabulary configurations in which there is a preponderance of irregulars, in early training, the adoption of a default regularization process is blocked. However, once the proportion of regular verbs exceeds 50%, the absolute number of regulars in the training set are allowed to act in consortium to produce generalization effects. The clear implication of this finding for predicting the onset of a default regularization process in children (such as overregularization errors) is that careful attention must be paid to the composition of children's verb vocabularies. These results *do not* indicate that an unrealistic imbalance in the input in favor of regular items is required for a connectionist account of the onset of a default regularization process, as Marcus et al (1992) have claimed.

Network Resources

In this final section, we evaluate the role of computational resources in predicting the networks' tendency to generalize by varying the number of hidden units included in each network. Nine sets of simulations were conducted in which the number of hidden units varied from 10 to 50. For any given set of simulations the number of hidden units was held constant throughout, i.e., at 10, 15, 20, 25, 30, 35, 40, 45 or 50. In all other respects, these simulations were identical in training schedule and vocabulary configuration to those described above. Recall that the standard networks used a total of 30 hidden units.

Table 5 summarizes the final generalization rates to novel indeterminate stems after vocabulary size has reached 500 verbs for each of the nine hidden unit conditions. In those simulations in which 20 hidden units or less were used, generalization to indeterminate novel stems remained consistently poor (<50%) throughout training. Performance when a larger number of hidden units (i.e., 25 to 50) were used resulted in no significant changes in generalization during the early training periods. However, after vocabularies exceeded 60 items, some differences were observed. For example, in simulations with 25 hidden units, generalization increased as would be expected as a function of vocabulary size early in training, but flattened out at around 50% beyond the 60 verb mark. In contrast, simulations in which 30–50 hidden units were used demonstrated continuously increasing generalization. Each of these simulations were indistinguishable from the pattern of performance outlined in Figure 4a. Within this hidden unit range (30-50 units), no systematic differences were observed regarding the relationship between the number of hidden units and the level of generalization ability. This result is not entirely surprising. It is a common finding that generalization in multi-layered networks is sensitive to the number of hidden units available. When hidden unit numbers are low (relative to problem size), generalization increases rapidly as hidden unit numbers are increased. There is then a range of hidden units where generalization remains roughly constant. However, as hidden unit numbers are further increased, generalization *decreases* but at a slow rate (Hinton, 1989). The generalization curve thus resembles a landscape (Occam's Hill) where one side is steep (the side with few hidden units) and the other side is a gradual incline (the side

with many hidden units). These control simulations show where the peak of Occam's Hill lies for problems of this type.

Together with the results reported in the previous section on vocabulary structure, these results suggest that within certain resource limits, changes in the generalization characteristics of the network are best attributed to reorganizations in the network resulting from qualitative and quantitative changes in the training set, as opposed to factors associated with pre-wired architectural considerations. The network appears to require sufficient representational resources to do its job (learn the task) but once those resource requirements are met, generalization properties are best understood by reference to characteristics of the training set.

Discussion

The major goal of this work was to provide a sketch of the mechanisms that might trigger the transition from rote learning to system building in children acquiring the English past tense. Like children, artificial neural networks can be observed to undergo several changes in performance over the course of mastering this system of inflectional morphology. These changes characterize the transition from rote learning to system building. In general, the results of the simulations indicated that given specified architectural constraints, this transition is not triggered simply by continued exposure to a subset of forms. Instead, the system requires gradual and incremental changes in the quantity of verbs which share specific mapping properties and characteristic phonological features that predict class membership.

Early in training after the initial set of 20 verbs are learned to criterion, the network did not exhibit any evidence of systematic rule-like behavior. Most newly presented regular and irregular verbs were incorrectly mapped (see Figure 2) and errors produced by the network during this period were not easily classifiable, i.e., they did not reflect the mapping characteristics of the initial learning set (see error tabulations in Table 3). Similarly, responses to novel verbs were unsystematic early in training, although some sensitivity was observed for novel forms which shared phonological characteristics with the irregular verbs in the initial training set (e.g., dental final consonant and vowel-consonant cluster). These results indicate that the network has failed to extract any pattern of regularities that can be said to define a default mapping strategy. Each regular stem/past tense pair is treated as an independent mapping problem. In essence, the network has *memorized* the initial set of regular stem/past tense pairs.

Nevertheless, the network's treatment of novel stems which share phonological features with the irregular stems in the initial training set indicates that the initial representations built up by the network cannot be defined entirely in terms of rote processes. Not only do the frequency characteristics of irregular verbs during early training contribute to early learning but their phonological coherence promotes faster learning¹⁰ and a limited degree of generalization to

¹⁰ Identity and vowel change verbs were the first verbs to be successfully mastered in the initial training set.

other forms. This result predicts that phonological factors, as well as frequency effects, will have an impact on children's early learning of irregular verbs.

The overall ability of the network to assimilate correctly new verbs in the training set (particularly on regular verbs) remains poor as vocabulary size increases (see Figure 2). However, poor performance on additional verbs is reversed when the vocabulary size exceeds 50 verbs. Beyond this point in training, performance on the new verbs improves rapidly. These findings indicate that the network is beginning to restructure its representation of the mapping problem, to the extent that the network now treats new verbs entering the vocabulary in a systematic fashion.

The comparison of performance in networks using various initial vocabulary conditions (see Table 4) suggests that the trigger for the transition from rote learning to system building in these networks is associated with the *quantity* of verbs in the training set which undergo systematic mapping processes. With respect to this vocabulary, the prime instigator of generalization (and hence internal reorganization in the network) was the number of *regular verbs* included in the initial vocabulary. Interestingly, the generalization of the suffix to novel stems can be seen to be blocked when irregular verbs constitute a majority of the training set. However, generalization takes off when regular stems constitute a majority in the training set and the absolute number of regular stems has reached a minimum threshold (approximately 50 verbs). Lastly, extensive training on limited vocabulary sets is insufficient to promote generalization (see Figure 5).

Taken together, these results demonstrate striking similarities between connectionist networks and children in the developmental course of learning inflectional morphological systems like the English past tense. To this extent, these models suggest that a *single mechanism* learning system may offer an alternative account of the transition from rote learning processes to system building in children's acquisition of English verb morphology. In contrast to the traditional view which posits an interaction between two qualitatively distinct mechanisms supporting different modes of representation (an associative, lexically based system for the irregulars, and a rule-based system for the regulars) (Pinker, 1991; Pinker & Prince, 1988; Marcus et al., 1992), a connectionist account posits a single mechanism driven by a general

learning algorithm which is capable of both memorization and generalization processes. The network's transition from rote-like to rule-like processes does not reflect the onset in operation of a separate rule-based mechanism. Instead, the shift in strategy marks an internally generated reorganization, triggered by *quantitative* increments in the size of a structured training set.

The behavior of these networks can be seen to mimic several aspects of the type and timing of children's pattern of morphological acquisition. The clearest parallels between network performance and English children's acquisition of inflectional verb morphology include:

- (1) An initial stage of acquisition where no overregularizations are observed.
- (2) A protracted period of acquisition where low rates of overregularization are observed (typically < 10%) but which gradually fade away as experience with irregular forms increases.
- (3) A final stage where overregularizations are not observed and the past tense forms of both regular and irregular verbs are produced correctly.

Evidence for (1) and (2) is well-documented in the literature (Bybee & Slobin, 1982; Marchman, 1988; Marcus et al., 1992). In particular, it is now well-accepted that overregularization errors occur selectively with a small proportion of irregular verbs in children's vocabularies, rather than in a system-wide fashion where all irregular verbs are susceptible to errors. However, the evidence for (3) is less clear-cut. Even adults sometimes produce regularization, and to a greater extent, irregularization errors (Bybee & Slobin 1982; Bybee & Moder, 1983; Stemberger & MacWhinney, 1986; Marchman & Plunkett, 1991). It will be left to future studies to assess the validity of the predictions of a single versus a dual mechanism system regarding the degree to which systems of inflectional morphology are susceptible to erroneous usage in older children and adults (see also Marchman & Plunkett, 1991, in preparation).

Additional parallels can be gleaned from the recent detailed analyses of longitudinal databases of American English-speaking children by Marcus et al. (1992). These include:

- (4) Irregular verbs with a high token frequency (e.g., *arbitrariness*) are less prone to overregularization than irregular verbs with a low token frequency (e.g., *identity* and *vowel change*).

- (5) The rate of overregularization of irregular forms is substantially less than the rate of regularization of regular stems.
- (6) On occasion, children, and in certain circumstances, adults will irregularize regular and irregular verbs.

Evidence for (4) is well-established (Slobin, 1971; MacWhinney, 1978), but has been documented in more detail in the recent work by Marcus et al (1992). Marcus et al (1992) also provide evidence for (5) (pg. 56-59) and use the finding to argue for the view that overregularizations reflect a tendency by the child to discriminate between regular and irregular verbs. In the current context, however, it should be noted that this behavioral distinction between regular and irregular verbs cannot be used to infer differentiation in the mechanisms that control the production of regular and irregular verbs. The same single mechanism is responsible for the distinctive regularization rates. Evidence for (6) is provided by Marcus et al's (1992) analysis of spontaneous speech corpora for Adam, Eve and Sarah. Bybee and Slobin (1982), Marchman (1988), and Marchman & Plunkett (1991) provide irregularization data obtained from elicitation tasks using both children and adults. In addition, Bowerman (personal communication) notes that her daughter Eva irregularized 6 irregular verbs and one regular form (*pick*→*puck*) while Christy irregularized 4 irregular verbs but no regular verbs. Marcus et al. (1992) also report that irregularizations of irregular verbs (though rare) are more frequent than irregularizations of regular verbs (page 33). In addition, the inappropriate irregularization of irregular verbs seemed to be conditioned by the no-change subregularity of a dental-final consonant (page 59,124).

Finally, consider the relationship between vocabulary size and onset of overregularizations predicted by these series of simulations. In both the experimental and control simulations, it was demonstrated that the generalization characteristics of the network (which are also responsible for the occurrence of overregularization errors) are related to the number of regular verbs in the training set, rather than continued training. However, recall that the first overregularizations are observed when vocabulary size exceeded approximately 110 verbs (see Figure 3). Further, the rate of increase in the tendency to generalize to indeterminate novel stems flattened off at around 140 verbs (see Figure 4). Yet, this is not the period in training in which the most drastic changes in vocabulary composition are observed. In this

vocabulary range, regular tokens comprised more than two-thirds of the verbs in the training set (approximately 71%) and the rate of increase in proportion of regular verbs in the training set had already flattened off. (Recall Figure 1 which shows that the fastest increases in regular verb vocabulary occurred when total vocabulary is less than 100 verbs.) Hence, it may seem odd that the onset of overregularizations (and the flattening off the generalization curve) does not occur earlier, at a time more closely yoked to drastic changes in vocabulary composition. The explanation for the fact that changes in response characteristics are not observed immediately is that modifications to the weight matrix of the network are made gradually. The full effect of training on any particular verb builds up gradually as the network is repeatedly exposed to the verb. The learning algorithm used by the network thus incorporates a built in delay (the learning rate) such that exposure to any given verb will not have its full effect on the response characteristics of the network until later in training. In addition, the proportion of regular verbs in the vocabulary does continue to increase (though at a slower rate) throughout training.

There has been little empirical investigation of the relationship between vocabulary size and overregularization reported in the child language literature. The findings that do exist are apparently contradictory. Marchman & Bates (in press) report a significant non-linear relationship between vocabulary size (types) and the proportion of verb vocabulary that is overregularized. In contrast, Marcus et al. (1992) find no relationship between increases in vocabulary size and the onset of overregularization errors. Comparison of these two studies is complicated by the fact that different data collection procedures were used. Marchman & Bates (in press) use parental report as their source; whereas, Marcus et al. (1992) analyze transcripts of spontaneous speech. Further, the ways in which overregularization rates are measured are also different in the two studies.

We suggest that the results from our simulations may serve to reconcile these diverse findings. In Figure 6a, we plot the proportion of regular verbs in the total vocabulary and proportion of overgeneralization errors for Adam as outlined in Marcus et al. (1992, their Figure 28, pg. 92). In Figure 6b, we plot the same data for the simulations reported above.

Insert Figures 6a and 6b about here

There are several striking similarities between the two figures. Looking first at Adam, note that the onset of overregularization errors occurs when the proportion of regular verbs in total vocabulary varies between 69% and 74%. In fact, the proportions of regular verbs varied between 67% and 77% during the entire period in which overregularization errors were reported for Adam by Marcus et al. (1992). In the simulation data (Figure 6b), overregularization errors are first observed when regular verb tokens make up a similar proportion of the total vocabulary, approximately 70%. For the entire period during which overregularizations occur in the simulations, regular verb proportions vary from 70% to 77%. Thus, the onset of overregularization errors is observed given similar vocabulary configuration conditions for both Adam and the simulations, i.e., when proportions of regular verbs are confined to a fairly constant narrow range. In general then, the data for Adam as well as the simulations reported here do not demonstrate a relationship between changes in vocabulary composition and overgeneralization rate *after vocabulary size has achieved a particular level*. In a sense, this should come as no surprise as far as the simulations are concerned given that earlier work has shown (Plunkett & Marchman 1991) that overregularization errors will occur even when vocabulary size and composition are *constant*.

Nevertheless, we know from the control simulations described earlier that the default regularization process (and hence overregularization errors) does not develop in the network unless the proportion and total number of regular verbs in the training set is sufficiently large. In terms of the simulation run shown in Figure 6b, the major shift in proportion of regular verbs occurs as vocabulary size expands from 20 to 60 verbs (i.e. from 42% regulars to 67% regulars). This shift creates the necessary input conditions for overgeneralization errors. However, such a change in input conditions is not sufficient for the onset of such errors -- the network itself must have time to *respond* to the changing input conditions. The representation of a default regularization process within the weight matrix of the network does not emerge all at once in its entirety; instead, the series of small adjustments to the connections whenever an output discrepancy is detected allows the network to construct a representation of a default mapping over time. As the connections strengthen in favor of a default regularization process,

the conflict with the irregular verbs in the training set becomes apparent and overregularization errors occur.

For Adam (Figure 6a), we observe no consistent switch in the balance between regular verbs and irregular verbs, and yet sporadic overregularization errors occur. These observations are consistent with two alternative interpretations:

- (1) There is no relation between the proportion and number of regular verbs in Adam's vocabulary and the onset of overregularization errors, as Marcus et al. (1992) claim.
- (2) The onset of overregularization errors is triggered by, but delayed with respect to, the achievement of an appropriate proportion and number of verbs in Adam's vocabulary.

Note that the characterization offered in (2) essentially yields the same type of explanation as that given for the onset of overregularization errors in the network. However, in order to substantiate the plausibility of the second interpretation, we must demonstrate that overregularizations do not occur within a period in Adam's development where the number and proportion of regular verbs in his vocabulary is *insufficient* to support a default regularization process. In effect, this means searching for overregularization errors in a period of development where regular verbs constitute less than 50% of total verb vocabulary and where there are less than 75 verbs. Unfortunately, the data reported by Marcus et al. (1992) for Adam, and indeed all of the other children in their study, cover a period of development where numbers and proportions of regular verbs are already sufficient to support a default regularization process. It is, therefore, not possible to evaluate the second interpretation of the data.

In contrast, data collected by Marchman & Bates (in press) permit an evaluation of the production of overregularization errors in children whose vocabularies contain as few as 0 to more than 600 verbs. Summarizing their findings, the parental report data in Marchman & Bates (in press) indicate that overregularization errors rarely are reported for children with small verb vocabularies (less than, say, 40 verbs), yet, there is a non-linear relation between verb vocabulary size and the proportion of those items that are subject to overregularization as verb vocabulary increases, i.e., rate of overregularization is not a simple linear proportion of total verbs produced by the child but exhibits a non-linear mass action effect. We suggest that the apparently contradictory findings reached by Marcus et al. (1992) and Marchman & Bates (in press) can be resolved by pointing to the fact that the two studies calculate correlational statistics

over different ranges of vocabulary size for the children involved. Marchman & Bates' (in press) data span the period of verb vocabulary growth which marks the transition from rote learning to system building (if interpretation (2) is correct). Hence verb vocabulary size is seen to correlate with overregularization errors. Marcus et al.'s (1992) data spans only a period of development where verb vocabulary already supports a default regularization process. During this period, the default regularization process is gradually strengthened and interferes intermittently with the correct inflection of old and new irregular verbs. As shown in Plunkett & Marchman (1991), the occurrence of a single overregularization during these conditions of training depends on a complex interaction of the current strength of the regularization process, on the one hand, and the token frequency and phonological characteristics of irregular verbs on the other hand. Since irregular verbs may be more or less strongly represented in the linguistic system at a given point in development, overregularization errors will occur in a sporadic fashion. Hence, no correlation between regular verb vocabulary size and the rate of overregularization need necessarily be observed during this later period of vocabulary growth.

In general then, we can identify two periods in the training of the network which are consistent with the hypothesis that vocabulary size plays a substantive role in the learning of inflectional verb morphology in young children. An early period of training (less than 100 verbs) results in a rapid transition in generalization characteristics. During this period, the rate of generalization (as measured by rate of suffixation of indeterminate novel verbs) increases almost linearly with the number of regular verbs in the vocabulary. However, no overregularization errors are observed because the generalization tendency of the network is not sufficiently strong to overcome the patterns of mapping for the irregular verbs on which the network has been trained. In the second period (signaled by an flattening off in the rate of increase in proportion of regular verbs in the vocabulary), the generalization tendency is consolidated and the first overregularizations are observed. The occurrence of overregularizations is then determined by a complex interaction between the frequency of irregular verbs and the phonological patterns of similarity and coherence which both encourage and protect irregular verbs from interference from the default regular mapping represented by the network.

These dynamics of network performance are entirely consistent with the absence of early overregularization errors and the non-linear relation between the size of regular verb vocabulary and the reported overregularizations for the children analyzed by Marchman & Bates (in press). Indeed, Marchman & Bates were originally prompted to examine this hypothesis based on predictions arising from earlier versions of the simulations described in this paper. However, we claim also that the data reported by Marcus et al. (1992) are consistent with the network facts described above, although these researchers give them their own interpretation. It remains to be shown that the mechanism which gives rise to the delayed onset of overregularization errors in the network, i.e., the gradual strengthening of an emergent default regularization process, is the same type of mechanism that is responsible for the onset of overregularization errors in children. Nevertheless, we can conclude that the available data are compatible with the view that overregularization errors in children may arise from a process of mass action (i.e., achievement of a critical mass of regular verbs) in a single mechanism system. In the single mechanism system, the passage from a period where internal representations are dominated by rote learning processes to a period of system building in which the system is capable of (over)generalizing novel stems and trained stems is driven by the gradual and incremental exposure to regular verbs in the training set.

One of the main findings of this work has been that a default regularization process will emerge in a network when it is exposed to a suitably configured vocabulary for a sufficient number of training trials. The default regularization process is blocked when the network recognizes verb stems which it has been taught are irregular. Two factors enable irregular verb stems to block the default regularization process: (a) their high token frequency, and (b) their phonological similarity to other irregular verbs. In the English past tense, irregular verbs are in the minority. They have a low type frequency as compared to the very large numbers of different regular verbs. Indeed, it is precisely the low type frequency of irregular verbs rather than their mapping characteristics which make them irregular in English. However, it is not always the case that the default mapping process coincides with the mapping characteristics of the largest class in the inflectional system. For example, the Arabic plural system is dominated by a large number of irregular mapping processes with only a relatively few forms undergoing

the default process. In the context of the present study, inflectional systems such as these would appear to pose problems for an approach that advocated the emergence of a default regularization process as the result of a process of mass action. However, Plunkett & Marchman (1991) have argued that a minority default regularization process can emerge in a neural network when the irregular mappings are contingent upon the presence of a defining feature in the stems that undergo the irregular mappings. In other words, the network can learn to apply a default mapping when the stem *lacks* features associated with other mapping types. In English, there are no phonological features that absolutely define an irregular past tense class (Pinker & Prince 1988, Plunkett & Marchman 1991). In the case of the Arabic broken plural, in contrast, all irregular stems possess phonological features that are clear predictors regarding the type of irregular mapping that is required (McCarthy & Prince 1990). In the absence of any of these features, the default mapping applies. Hare & Elman (1992) have made a similar point with respect to understanding diachronic shifts in the configurations of class membership in the inflectional systems of Old English.

Not all inflectional systems with a minority default regularization process possess irregular mappings which are triggered by the presence of defining phonological features. For example, Clahsen, Rothweiler, Woest & Marcus (1992) describe the different mapping types associated with the German plural system. The German plural system differs from English in that it does not possess a single inflectional class that comprises the majority of nouns in the language. They also point out that there are no defining characteristics associated with the mapping types – only loose tendencies can be seen to characterize a given inflectional class. Furthermore, Clahsen et al (1992) report that young German children treat the /-s/ and/or /-en/ endings as default plurals, i.e., overregularization errors are always of these inflectional types, even though neither form has a substantial representation in their vocabularies. At first blush, these findings might be seen to embarrass an account that relies exclusively on the achievement of a critical mass for the emergence of a default mapping process. However, as we have noted above, there are multiple determinants that need to be considered when predicting the inflectional mapping of a stem in a neural network. Although there are apparently no defining characteristics associated with the different inflectional classes of the German plural, weak

distributional tendencies associating phonological features within a class and across classes might act in consortium to support learning and representation of a minority default process by a neural network. For example, MacWhinney et al. (1989) showed how a neural network could exploit such weak distributional information in its categorization of German nouns according to gender. Even if such weak phonological information proves inadequate to achieve appropriate inflectional categorization, other sources of information might substitute or supplement the categorization of the German plural. Clahsen et al. (1992) note that the /-en/ form is used predominantly with female gender nouns. This sub-regularity may serve to encourage overregularization errors of the /-en/ type and/or act in consortium with other non-definitional features to help partition the inflectional space of the German plural. Indeed, Wurzel (1989) points to a complex interaction between gender and phonological characteristics for the determination of class membership in the German plural system.

Note, however, that by introducing gender information to the network, we are introducing a new representational resource to the input system. Indeed, any non-phonological information that might be required to resolve the partitioning of the inflectional space of a language constitutes an adaptation to the models we have discussed thus far. There are good reasons to believe that such adaptations are necessary. For example, Pinker & Prince (1988) point out that a network that relies exclusively on phonological information at the input has no principled resource to resolve the multiple mappings of homophonic stems such as /ring/. Most of the connectionist verb learning models have relied exclusively on phonological information in making their behavioral predictions. Indeed their ability to predict much of what is known about children's acquisition of the English past tense, based solely on networks that map from a phonological representation of the stem to a phonological representation of the past tense, has been an important contribution to our understanding of the developmental process. However, as MacWhinney & Leinbach (1991) point out: "Implementations are not conceptualizations." If a network model can benefit from supplementary sources of information and, at the same time, provide additional insights into the acquisition process, then adaptations or reformulations of existing network frameworks must be welcomed. For example, recent work by Cottrell & Plunkett (1991) has shown how the homophone problem, noted above, can be resolved in a

network that maps from semantics to phonology in a fashion that does not rely on semantic similarity to predict inflectional type (as it must not – see Pinker & Prince (1988) on the *hit, strike, slap* problem). Interestingly, these reformulations of the past tense problem permit researchers to address other issues which are beyond the scope of simple feedforward phonological-to-phonological networks. For example, it becomes more plausible to construct networks that attempt to map entire inflectional systems including both noun and verb inflections. Networks that perform multi-class mappings of this type can then be evaluated against the known facts of allomorphic variation which span inflectional systems or the inflectional properties of words that migrate to other classes, such as denominal verbs. It remains to be seen whether extensions of this type are necessary for explaining phenomena like the minority default process observed in German children by Clahsen et al. (1992). The limitations of the representational resources to which a neural network is exposed during training is a powerful source of hypotheses as to the nature of the information children may exploit in their construction of the inflectional system.

Conclusion

These series of simulations sought to provide further understanding of the acquisition of the regularities and exceptions that are inherent in many linguistic systems, in particular, the English past tense. Our current understanding about children's acquisition of English inflectional morphology is typically based on naturalistic observations of children's actual productions, and experimental studies using nonsense words (e.g., Berko, 1958). It is worth noting that these two types of data have generated seemingly contradictory findings regarding the nature and frequency of past tense errors. In particular, naturalistic studies report that overgeneralization errors comprise a small portion of children's productive vocabularies, though children do produce a range of regularization and, in isolated cases, irregularization errors. In contrast, studies using nonsense forms typically demonstrate an overwhelming preference for regularization. Several studies have suggested that children are sensitive to the phonological aspects of nonsense forms. For example, Derwing & Baker (1986) have shown that novel

stems which end in a dental tend to be identity mapped and novel stems with specific vowel/consonant clusters will sometimes undergo vowel suppletion.

In these simulations, analyses of network performance on trained verbs is analogous to naturalistic observations of children's performance on real English words. Correspondingly, network performance on novel verbs is best viewed to correspond to the experimental elicited production studies with nonsense verbs. Interestingly, the overall pattern of results suggests a considerable correspondence between children and networks across these two types of measures. For example, the pattern of errors observed across learning for verbs in the training set are multi-lateral, i.e., irregular verbs are regularized and regular verbs are irregularized. In contrast, responses to novel verbs are overwhelmingly more likely to involve regularization. For those novel stems that phonologically resemble irregular verbs, the tendency to regularize is still prominent but identity mappings and vowel suppletions do occur. Thus, evaluations of network performance on novel verbs (in comparison to patterns of errors across learning) suggests that the network, like the child, is best characterized as a rule-governed system.

We suggest that evaluations of acquisition with experimental studies using nonsense verbs, while clearly illustrating the generalization abilities of young children, may have succeeded in biasing our view of the nature of the phenomenon to be explained. While we accept that experimental findings contribute important insights into the developmental process, they do not obviate the need for detailed naturalistic studies of children's acquisition of verb morphology. Further analyses of longitudinal databases are required to evaluate theories concerning the acquisition of inflectional systems in young children and in particular the role of vocabulary size for the process of building systematic linguistic knowledge representations.

References

- Bates, E., Bretherton, I., & Snyder, L. (1988). From First Words to Grammar: Individual Differences and Dissociable Mechanisms. Cambridge, MA: Cambridge University Press.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J. & Hartung, J. (1992). Developmental and stylistic variation in the composition of early vocabulary. CRL Technical Report # 9205, University of California, San Diego.
- Berko, J. (1958). The child's learning of English morphology. Word, 14, 150-177.
- Bever, T. G. (1982). Regressions in mental development: Basic phenomena and theories. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brown, R. (1973). A first language: The early stages. London: Allen & Unwin.
- Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. In E. H. Lenneberg (Eds.), New directions in the study of language Cambridge, MA: MIT Press.
- Bybee, J., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. Language, 58, 265-289.
- Bybee, J., & Moder, C. L. (1983). Morphological classes as natural categories. Language, 59, 251-270.
- Carey, S. (1982). Semantic Development: The state of the art. In E. Wanner & L. Gleitman (Eds.), Language Acquisition: The State of the Art Cambridge, MA: Cambridge University Press.
- Cazden, C. B. (1968). The acquisition of noun and verb inflections. Child Development, 39, 433-448.
- Clahsen, H., Rothweiler, M., Woest, A. & Marcus, G. F. (1992) Regular and irregular inflections in the acquisition of German noun plurals. Cognition, 45, 225-255.
- Cottrell, G. W. & Plunkett, K. (1991). Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sound. Proceedings of the thirteenth annual Cognitive Science Society Conference, Chicago, IL, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Derwing, B. L., & Baker, W. J. (1986). Assessing morphological development. In P. Fletcher & M. Garman (Eds.), Language acquisition: Studies in first language development. Second edition Cambridge: Cambridge University Press.
- de Villiers, J. G., & de Villiers, P. A. (1985). The Acquisition of English. In D. I. Slobin (Eds.), The Crosslinguistic Study of Language Acquisition (Vol 1) Hillsdale, NJ: Lawrence Erlbaum.
- Dromi, E. (1987). Early lexical development. New York, NY: Cambridge University Press.

- Ervin, S. M., & Miller, W. R. (1963). Language development. In H. W. Stevenson (Eds.), Child psychology: The Sixty-Second Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press.
- Elman, J. (1991). Incremental learning, or the importance of starting small. CRL Technical Report No. 9101, University of California, San Diego.
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (In press). MacArthur Communicative Development Inventories: User's Guide and Technical Manual, San Diego, CA: Singular Press.
- Hare, M. & Elman, J. L. (1992). A connectionist account of English inflectional morphology: Evidence from language change. Proceedings of the fourteenth annual Conference of the Cognitive Science Society, Bloomington, IN, Hillsdale, NJ, Lawrence Erlbaum Associates.
- Hinton, G. E. (1989). Connectionist learning procedures. Artificial Intelligence, 40, 185–234.
- Johnston, J. & Slobin, D. I. (1979). The acquisition of locative expressions in English, Italian, Serbo-Croatian, and Turkish. Journal of Child Language, 6, 529-545.
- Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. Cognition, 23, 95-147.
- MacWhinney, B. (1978). The acquisition of morphophonology. Monographs of the Society for Research in Child Development, 43(174).
- MacWhinney, B., Leinbach, J., Taraban, R. & McDonald, J. (1989) Language learning: Cues or rules? Journal of Memory and Language, 28, 255–277.
- MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb leaning model. Cognition, 40, 121–157.
- McCarthy, J. J., & Prince, A. S. (1990). Foot and word in prosodic morphology: The Arabic broken plural. Natural Language and Linguistic Theory, 8, 209–283.
- Maratsos, M. P. (1983). Some current issues in the study of the acquisition of grammar. In P. Mussen (ed.), Carmichael's Handbook of Child Psychology, 4th Ed. J. H. Flavell & E. H. Markman (Eds.), Vol. III: Cognitive Development. New York: John Wiley & Sons.
- Marchman, V. (1988). Rules and regularities in the acquisition of the English past tense. Center for Research in Language Newsletter, 2(4).
- Marchman, V., & Bates, E. (in press). Continuity in lexical and morphological development: A test of the critical mass hypothesis. Journal of Child Language
- Marchman, V. & Plunkett, K. (in preparation). When does an adult behave like a child?: On-line and naturalistic studies of the production of English inflected forms.

- Marchman, V., & Plunkett, K. (1991). Irregularization. Paper presented at the 16th Annual Boston University Conference on Child Language Development. Boston, Massachusetts, October 1991.
- Marcus, G. F., Ullman, M., Pinker, S., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. Monographs of the Society for Research in Child Development, 57(4), Serial No. 228.
- Pinker, S. (1991). Rules of Language. Science, 253, 530–535.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition, 28, 73-193.
- Plunkett, K. (1993) Lexical segmentation and vocabulary growth in early language acquisition. Journal of Child Language, 20, 43-60.
- Plunkett, K., & Marchman, V. (1989). Pattern association in a back propagation network: Implications for language acquisition TR No. 8902. Center for Research in Language, University of California, San Diego.
- 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.
- Prasada, S., & Pinker, S. (1993). Generalisation of Regular and Irregular Morphological Patterns. Language and Cognitive Processes, 8(1), 1–56.
- Rosenblatt, F. (1962). Principles of neural dynamics. New York: Spartan.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models (pp. 216-271). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations (pp. 318-362.). Cambridge, MA: MIT Press.
- Slobin, D. I. (1971). Developmental psycholinguistics. In W. O. Dingwall (Eds.), A survey of linguistic science College Park, Maryland: University of Maryland Press.
- Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Eds.), The cross-linguistic study of language acquisition Vol. 2: Theoretical issues Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stemberger, J. P., & MacWhinney, B. (1986). Form-Oriented Inflectional Errors in Language Processing. Cognitive Psychology, 18, 329–354.

Wurzel, W. U. (1989). Inflectional Morphology and Naturalness. Dordrecht: Kluwer Academic Press.

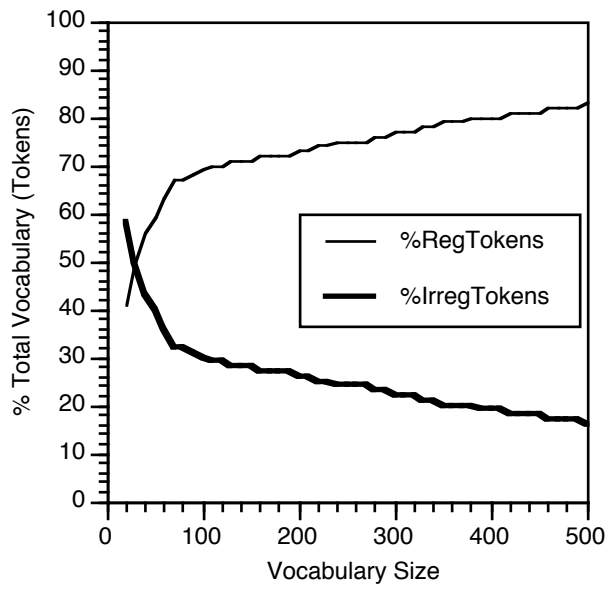


Figure 1: Proportion of regular and irregular verbs in training set as vocabulary size expands.

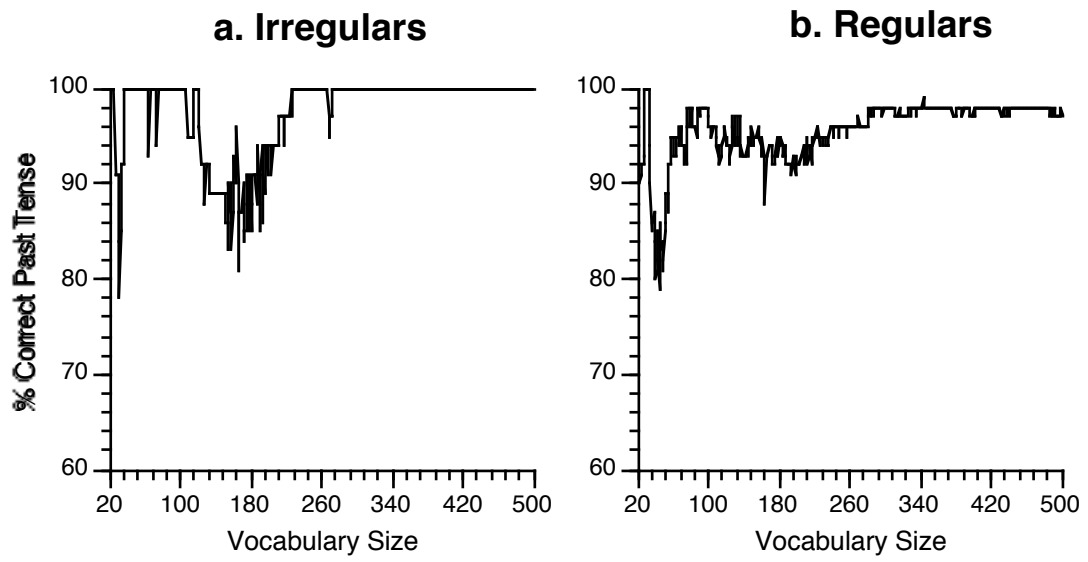


Figure 2: Hit rates for irregular and regular verbs

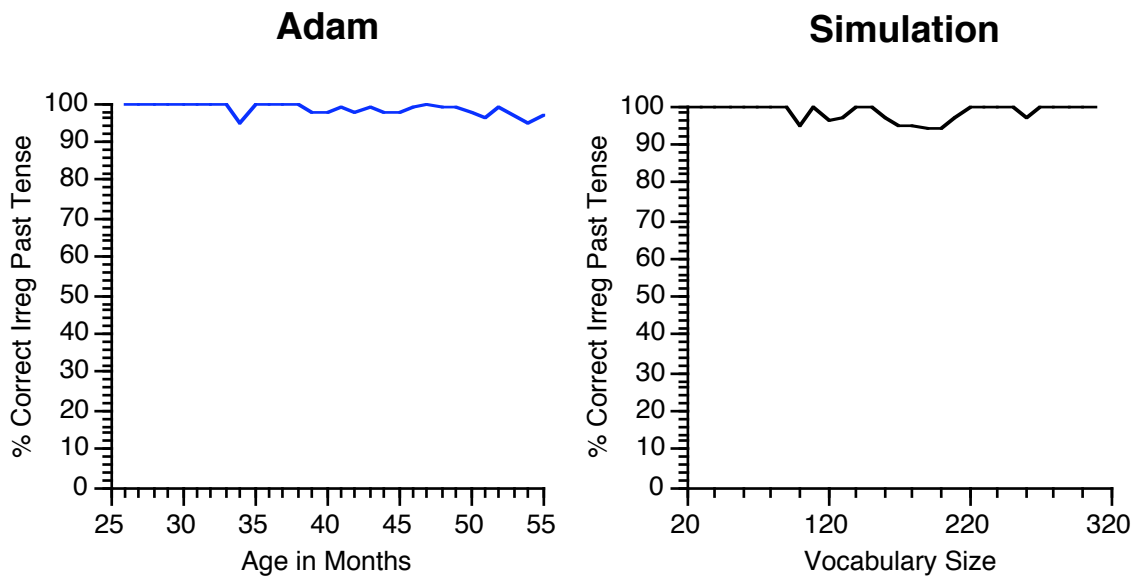


Figure 3: (1- Overregularization Rate) for Adam (reproduced from Marcus et al., 1992) and the simulation. Data are expressed as percentages of irregular verbs produced. The overregularization rate for Adam reflects the number of verb tokens; whereas, for the simulation it reflects the number of verb types (also see footnote 8).

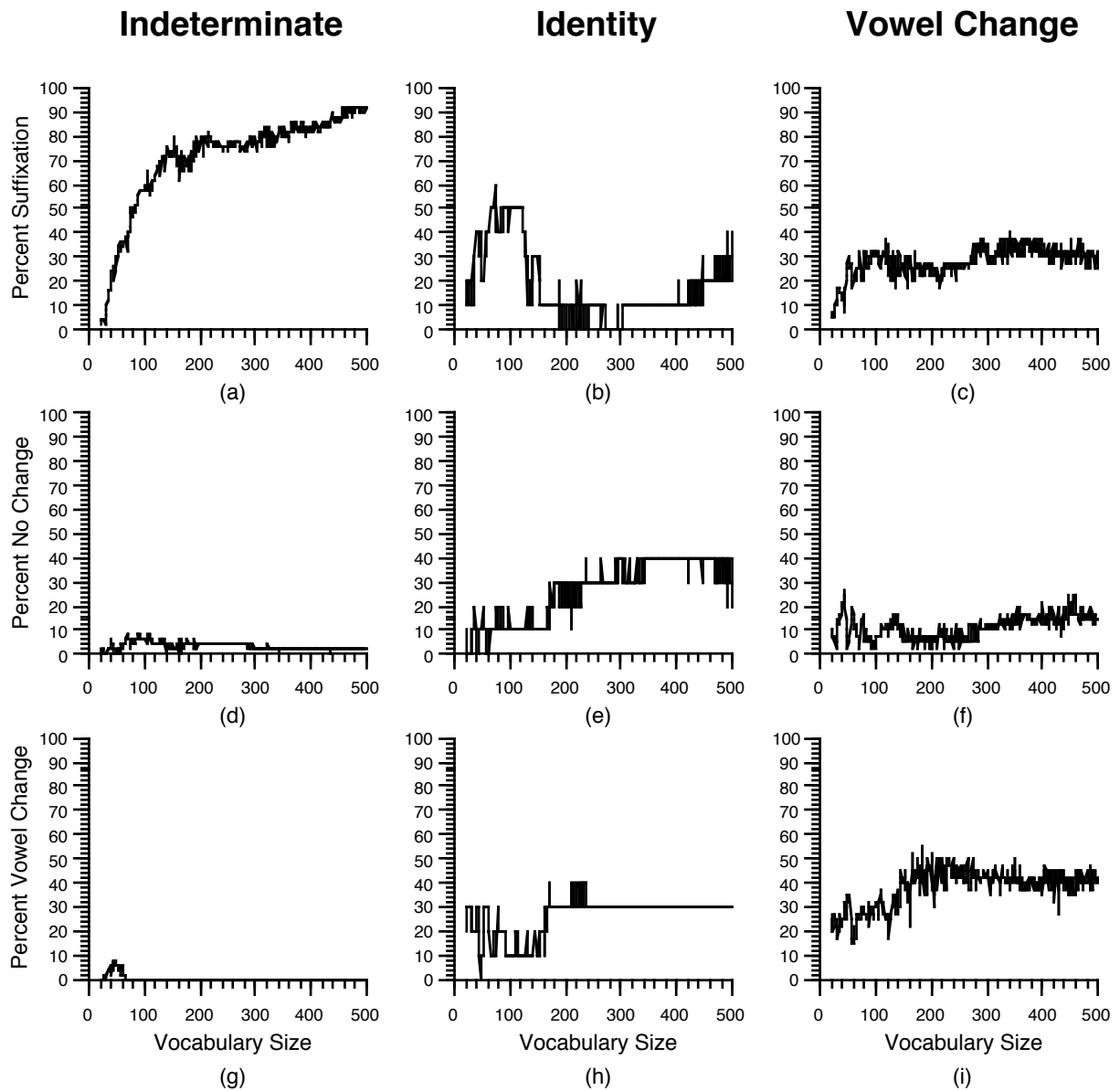


Figure 4: Network performance on Novel Verbs. Three types of novel stems are used; indeterminate stems which possess no defining characteristics; identity stems which possess a dental final consonant; vowel change stems which possess a vowel-consonant cluster characteristic of vowel change verbs in the training set. For each novel stem type, the percentage which are regularized, identity mapped or vowel changed is plotted.

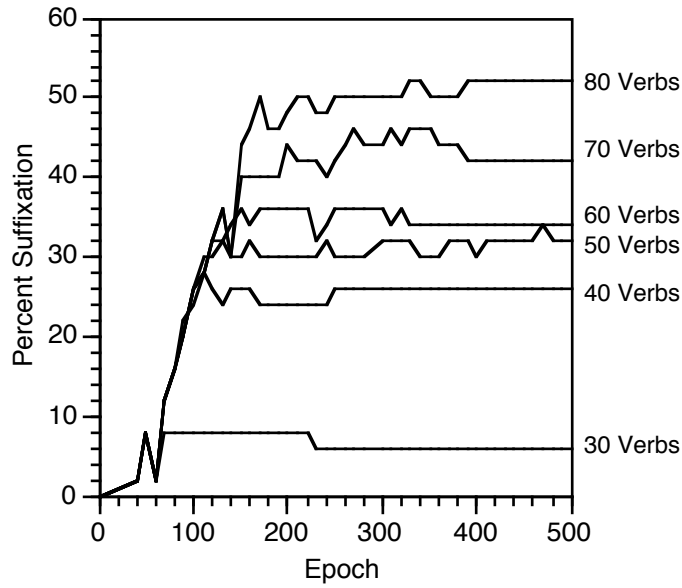


Figure 5: Generalization of Indeterminate Stems by vocabulary size. 6 sets of simulations are plotted. Each set corresponds to a different final vocabulary size (30, 40, 50, 60, 70 or 80 verbs). In each set, however, training continues for 500 epochs. Weight matrices are saved at different points in training and regularization of indeterminate novel stems evaluated.

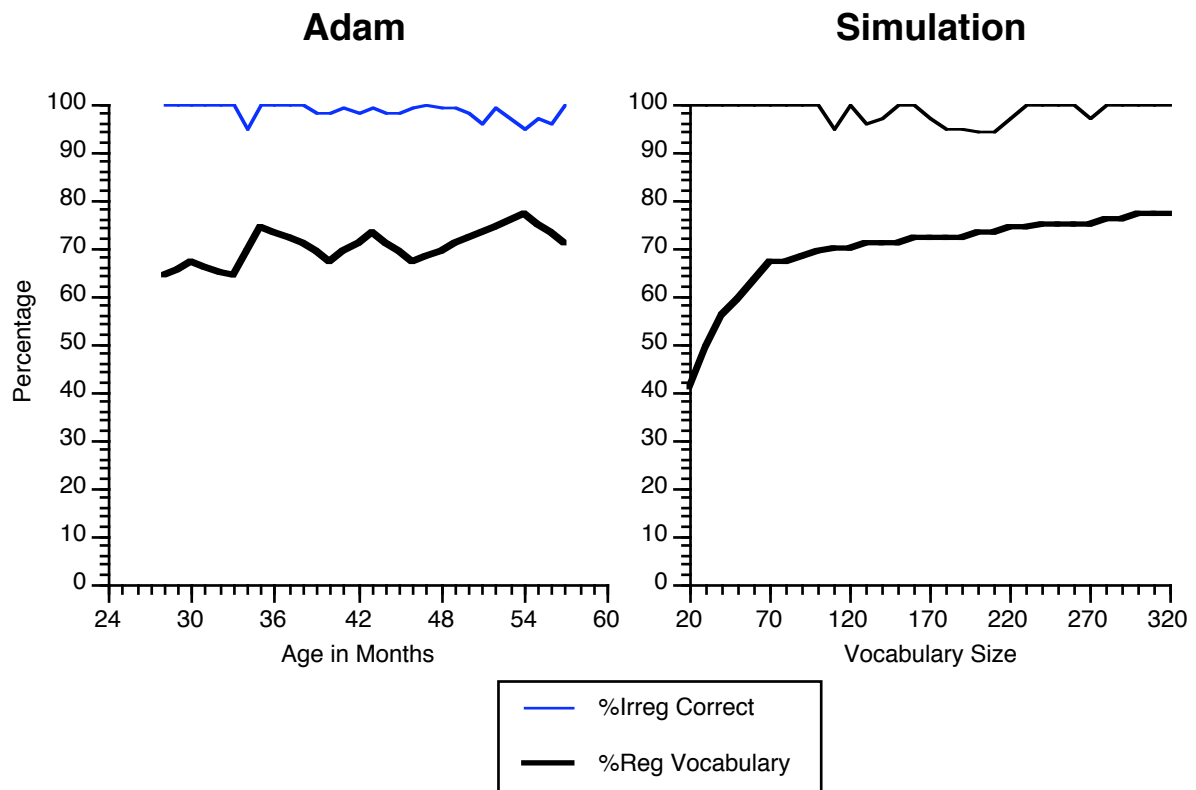


Figure 6: Compares overregularization rate and proportion of regular verbs of total vocabulary tokens for simulation with same measures from for Adam (taken from Marcus et al. (1992)). Note that the irregularization rate for the simulation reflects the number of irregular verb types (as in Figure 3b – see also footnote 8).

Table 1: Phonological Representation

		PHONOLOGICAL FEATURE UNITS					
		CON/VOW	VOICING	MANNER		PLACE	
		#1	#2	#3	#4	#5	#6
/b/		0	1	1	1	1	1
/p/		0	0	1	1	1	1
/d/		0	1	1	1	1	0
/t/		0	0	1	1	1	0
/g/		0	1	1	1	0	0
/k/		0	0	1	1	0	0
/v/		0	1	1	0	1	1
/f/		0	0	1	0	1	1
/m/		0	1	0	0	1	1
/n/		0	1	0	0	1	0
/ŋ/		0	1	0	0	0	0
/ð/		0	0	1	0	1	0
/θ/		0	1	1	0	1	0
/z/		0	1	1	0	0	1
/s/		0	0	1	0	0	1
/w/		0	1	0	1	1	1
/l/		0	1	0	1	1	0
/r/		0	1	0	1	0	1
/y/		0	1	0	1	0	0
/h/		0	0	0	1	0	0
/i/	(eat)	1	1	1	1	1	1
/ɪ/	(bit)	1	1	0	0	1	1
/o/	(boat)	1	1	1	0	1	1
/ʌ/	(but)	1	1	0	1	1	1
/u/	(boot)	1	1	1	1	0	1
/ʊ/	(book)	1	1	0	0	0	1
/e/	(bait)	1	1	1	1	1	0
/ɛ/	(bet)	1	1	0	0	1	0
/aɪ/	(bite)	1	1	1	0	0	0
/æ/	(bat)	1	1	0	1	0	0
/aʊ/	(cow)	1	1	1	1	0	0
/ɔ/	(or)	1	1	0	0	0	0

Table 2: Vocabulary Structure by Verb
Class

Total	Arbs	Regs	Ids	Vcs	Tok ¹¹
20	2	10	4	4	5
80	2	61	8	9	3
100	2	77	10	11	1
140	2	112	11	15	1
200	2	163	15	20	1
260	2	221	17	20	1
269	2	227	20	20	1
380	2	338	20	20	1
500	2	458	20	20	1

¹¹ Arbitrary verbs have a token frequency of 15.

Table 3: Hit rates and Errors by Verb Class. Performance is evaluated each time vocabulary size increases by 10 verbs. See text for interpretation of error coding for individual verb classes. All measures except for vocabulary size are percentages. The HIT columns provide the percentage hit rates for their respective classes. Error classification (SUF, ID, BLD, VC, UNC) are also percentage measures of all non-hits within a class and, therefore, sum to 100%. For the regulars, a measure of the proportion of types and tokens of total vocabulary is also provided.

VOC	REGULARS							IDENTITIES				VOWEL CHANGE					
	%TP	%TK	HIT	SUF	ID	BLD	UNC	HIT	SUF	VC	UNC	HIT	SUF	ID	BLD	VC	UNC
20	50	41	100	0	0	0	0	100	0	0	0	100	0	0	0	0	0
30	60	49	90	0	0	0	100	100	0	0	0	71	0	0	0	0	0
40	67	56	80	20	0	0	80	100	0	0	0	100	0	0	0	0	0
50	70	59	85	20	20	20	40	100	0	0	0	100	0	0	0	0	0
60	73	63	95	50	50	0	0	100	0	0	0	100	0	0	0	0	0
70	77	67	94	0	33	0	67	100	0	0	0	100	0	0	0	0	0
80	76	67	98	0	100	0	0	100	0	0	0	100	0	0	0	0	0
90	76	68	97	0	100	0	0	100	0	0	0	100	0	0	0	0	0
100	77	69	97	0	50	0	50	100	0	0	0	100	0	0	0	0	0
110	78	70	94	20	20	0	60	90	100	0	0	100	0	0	0	0	0
120	78	70	94	40	20	0	40	100	0	0	0	100	0	0	0	0	0
130	79	71	97	33	66	0	1	100	0	0	0	85	0	50	50	0	0
140	80	71	93	57	28	0	15	100	0	0	0	80	0	66	33	0	1
150	80	71	95	40	20	0	40	100	0	0	0	81	0	33	0	33	33
160	80	72	93	37	37	0	26	91	0	100	0	88	0	50	0	0	50
170	80	72	94	28	42	0	30	84	0	50	50	88	50	0	0	50	0
180	81	72	95	14	57	0	29	92	100	0	0	78	0	25	25	50	0
190	81	72	92	9	54	0	63	92	100	0	0	78	0	25	25	50	0
200	81	73	92	15	53	0	32	93	100	0	0	95	0	0	100	0	0
210	82	73	94	10	60	0	30	93	100	0	0	95	0	0	100	0	0
220	82	74	95	0	66	11	23	93	100	0	0	100	0	0	0	0	0
230	83	74	95	11	55	0	34	100	0	0	0	100	0	0	0	0	0
240	84	75	96	0	71	14	15	100	0	0	0	100	0	0	0	0	0
250	84	75	96	0	62	12	16	100	0	0	0	100	0	0	0	0	0
260	84	75	96	12	62	12	14	100	0	0	0	100	0	0	0	0	0
270	84	75	97	0	83	0	17	95	100	0	0	100	0	0	0	0	0
280	85	76	96	0	62	12	26	100	0	0	0	100	0	0	0	0	0
290	85	76	98	25	75	0	0	100	0	0	0	100	0	0	0	0	0
300	86	77	98	25	75	0	0	100	0	0	0	100	0	0	0	0	0
310	86	77	98	20	60	0	20	100	0	0	0	100	0	0	0	0	0
320	86	77	97	14	42	28	16	100	0	0	0	100	0	0	0	0	0
330	87	78	98	20	60	0	20	100	0	0	0	100	0	0	0	0	0
340	87	78	98	25	50	0	25	100	0	0	0	100	0	0	0	0	0
350	88	79	98	25	75	0	0	100	0	0	0	100	0	0	0	0	0
360	88	79	98	40	60	0	0	100	0	0	0	100	0	0	0	0	0
370	88	79	98	40	40	0	20	100	0	0	0	100	0	0	0	0	0
380	88	80	97	28	57	14	1	100	0	0	0	100	0	0	0	0	0
390	89	80	98	16	66	16	2	100	0	0	0	100	0	0	0	0	0
400	89	80	98	14	71	14	1	100	0	0	0	100	0	0	0	0	0
410	89	80	98	16	66	16	2	100	0	0	0	100	0	0	0	0	0
420	90	81	98	16	66	16	2	100	0	0	0	100	0	0	0	0	0
430	90	81	98	20	40	20	20	100	0	0	0	100	0	0	0	0	0
440	90	81	97	12	50	25	13	100	0	0	0	100	0	0	0	0	0
450	90	81	98	28	28	42	2	100	0	0	0	100	0	0	0	0	0
460	90	82	98	16	33	49	2	100	0	0	0	100	0	0	0	0	0
470	91	82	98	16	33	49	2	100	0	0	0	100	0	0	0	0	0
480	91	82	98	16	33	49	2	100	0	0	0	100	0	0	0	0	0
490	91	82	97	22	33	33	12	100	0	0	0	100	0	0	0	0	0
500	91	83	97	20	30	30	20	100	0	0	0	100	0	0	0	0	0

Table 4: Nine Starting Vocabulary Conditions: Vocabulary composition by regular and irregular verbs at the 20 and 40 vocabulary marks plus rate of generalization to novel indeterminate stems at the 40 vocabulary mark.

Cond	Vocabulary 20		Vocabulary 40		
	Regs	Irregs	Regs	Irregs	% Suffix
1	0 (0%)	20	0 (0%)	40	0
2	0 (0%)	20	5 (12.5%)	35	0
3	5 (25%)	15	10 (25%)	30	0
4	5 (25%)	15	15 (37.5%)	25	0
5	10 (50%)	10	20 (50%)	20	12
6	15 (75%)	5	25 (67.5%)	15	16
7	15 (75%)	5	30 (75%)	10	18
8	18 (90%)	2	35 (87.5%)	5	32
9	18 (90%)	2	38 (95%)	2	40

Table 5: Nine Hidden Unit Conditions: Generalization to indeterminate novel stems is measured at the end of training (500 verbs) in a series of simulations containing differing numbers of hidden units.

Condition	1	2	3	4	5	6	7	8	9
# Hidden Units	10	15	20	25	30	35	40	45	50
% Generalization	0	12	36	62	84	86	84	92	86