Acquiring the mapping from meaning to sounds

Garrison W. Cottrell[1]
Department of Computer Science and Engineering
University of California, San Diego

Kim Plunkett
Department of Experimental Psychology
University of Oxford

(to appear in *Connection Science*)

January 24, 1995

[1]We thank Steen Ladegaard Knudsen for his assistance in programming, analysis and running of simulations, Scott Baden for his assistance in vectorizing our code for the Cray Y-MP, the Division of Engineering Block Grant for time on the Cray at the San Diego Supercomputer Center, and the members of the PDPNLP and GURU Research Groups at UCSD for helpful comments on earlier versions of this work.

Abstract

One of the fundamental difficulties facing a child trying to acquire a language is that the association between meanings and sounds is for the most part an arbitrary one. In this work, we model this process using a recurrent neural network that is trained to map a set of plan vectors, representing meaning, to associated sequences of phonemes, representing the phonological structure of the surface forms. We evaluate the role of the similarity structure of the target forms (the adult vocabulary) and the similarity structure of the input forms (the semantic structure) on the evolution of the network's vocabulary. The model's performance offers a principled account of various phenomena associated with children's early vocabulary development including the difficulty of acquiring synonyms, the appearance of idiosyncratic forms and over-extension errors. The model makes several unexplored predictions for the developmental profiles of young children acquiring morphology.

1 Introduction

Most prior models of the acquisition of morphophonology in connectionist nets (MacWhinney & Leinbach [1991]; Plunkett & Marchman [1991]; Plunkett & Marchman [1993]; Rumelhart & McClelland [1986]) have focused on the problem of learning the relationships between phonological representations of various paradigms of verbs. That is, they take the problem to be one of learning an *intra*-level mapping between phonological forms. Second, earlier models have used feed-forward neural networks to model what is inherently a temporal process, that is, the generation of a phonological form. Our view is that the process of generation of phonological forms is inherently an *inter*-level process, one that involves a mapping from meanings to sounds. We assume that the generation process is a *temporal* process that proceeds in left-to-right fashion from the beginning to the end of the word. In this paper, we explore a particular model of this process, a connectionist network that generates one phoneme at a time from a fixed representation of the meaning of the word.

Our model is based on the work of Jordan [1986] and Elman [1990] (see Figure 1). We use a hidden-recurrent network to generate phonemes (represented as distinctive features) one at a time from a static input pattern (Jordan termed this the *plan vector*). Jordan investigated, for simple cases, the types of plan vectors that made learning sequences easy. In our case, we use input vectors that make learning hard, because we view the mapping being learned as one from meanings to motor programs. Thus, the input vectors have a similarity structure that is unrelated to the output sequences. We use a second layer between this input and the recurrent hidden layer so that the network may learn the appropriate plan vector for the sequence. The problem we set for the network is to generate the appropriate sequence of phonemic forms corresponding to the different morphological paradigms for a verb.

We are not the first to attempt a model of this type (Gasser & Lee [1990]; Hare [1990];

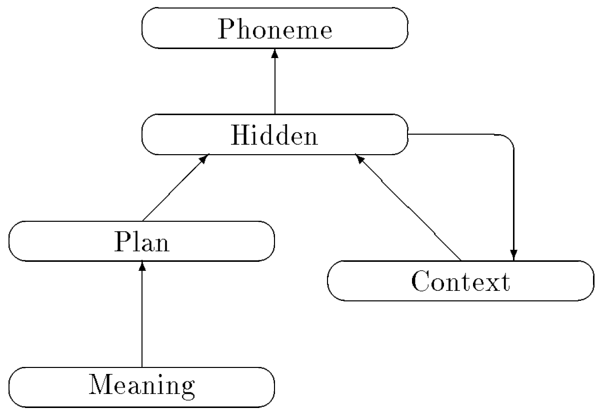


Figure 1: General network architecture: A meaning vector corresponding to the semantics of the present or past tense of the verb is held constant at the input layer while a sequence of phonemes is produced at the output layer. The context units permit the meaning vector to act as a plan for controlling the production of a sequence of phonemes in appropriate order. Phonemes are represented as vectors of distinctive features. Meanings are represented by two components; a lexical meaning plus the present vs. past tense. The plan layer has 50 units, the hidden and context 100, and the output layer 15 units. The input layer size varies with the representation used.

Hare, Corina & Cottrell [1989]; Jordan [1986]; Plaut & Shallice [1993]), however, we believe this is the first attempt to explore in a systematic fashion the effects of semantics on acquisition of surface forms. For example, Hare, Corina & Cottrell [1989] used a network which used localist (one unit on) meaning vectors. Since localist representations are inherently orthogonal, no similarity structure existed between different "meanings". In this paper, we use a representation of semantics as distortions of prototypes that imposes a *similarity structure* on the input to the network. Also, the target vocabulary has phonological regularities, resulting in a similarity structure in the output domain. These are associated at random. Thus the network is faced with a problem similar to that facing the child: There is a similarity structure in the input (presumably extracted from perceptual representations, and corresponding to objects and relations in the world) and there is a similarity structure in the targets (corresponding to the adult language forms), and there is a completely *arbitrary* relationship between the two.

Plaut & Shallice [1993], in a connectionist model of deep dyslexia, explored a similar problem, using a similar mapping structure (see below). However, where Plaut & Shallice were concerned with the effects of damage on a trained model, we are concerned with the learning trajectory, that is, the developmental aspects of the model. The Plaut & Shallice model used a small training set, which mitigated against a systematic evaluation of the role of semantic similarity in network performance. In contrast, we use a much larger training set, which permits this evaluation. The size of our lexicon also allows us to analyze the effects of surface form similarity in a more systematic way. Finally, Plaut & Shallice did not explore inflectional variation in their model.

We use a familiar domain to instantiate the model; the acquisition of the past tense in English. Thus the model will take as input a vector representation of the semantics of the verb, and produce as output the phonological representation. We assume this latter representation is a sequence of targets for a motor program.[1] In order for the model to produce different forms for the same verb, we include in our notion of semantics of the lexical item underlying semantic markers that distinguish surface phonological forms of present and past tense. This gives rise to a very different notion of the mapping from stems to past than the one that has been used previously (see Figure 2). The network of Rumelhart & McClelland [1986] for example, implemented an intra-level mapping between representations of the stem and past forms, corresponding to MAPPING 2 in Figure 2. Since we are implementing an inter-level mapping from meaning to form, corresponding to MAPPING 1 in Figure 2, MAPPING 2, that has been implemented directly by Rumelhart & McClelland is implemented *indirectly* by our network. That is, it is a relationship between different network outputs, rather than a relationship between network inputs and outputs.

In learning to perform this task, the network must exploit phonological regularities that

[1]To use these as specifications for a motor program, a forward model would need to be learned as outlined by Jordan & Rumelhart [1992].

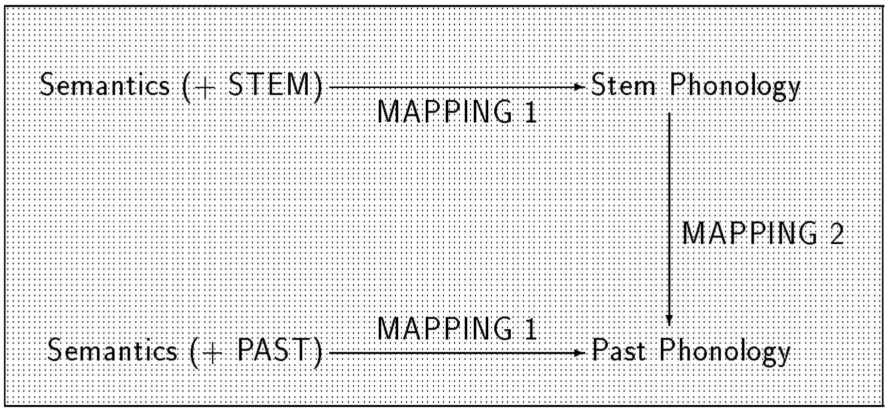


Figure 2: Two notions of "mapping": MAPPING 2 represents intra-level mappings between phonological representations or the stem and past tense forms of the verb. MAPPING 1 represents the inter-level mapping between the semantic representation of the verb and it various phonological forms. Under MAPPING 1 the relationship between the various phonological forms of the verb are expressed by the network in an indirect fashion.

characterize the indirect relation between stems and their past tense forms. In particular, in the regular paradigm, the choice of the ending is conditioned by the identity of the final phoneme of the stem.[2] For example, if the verb ends in a dental, the epenthesized form of the past tense (*/ed/*, as in *flitted*) is required. Such phonological information is exploited by children and adults when prompted for the past tense form of a novel stem (Bybee & Slobin [1982]; Marchman [1988]), and must be captured by our network in some way. However, the phonological characteristics of the verb stem do not uniquely *determine* its past tense form. Although all verbs which have identical stem and past tense forms possess a dental final consonant (e.g. *hit* → *hit*), not all verbs that end in a dental consonant have identical stem and past tense forms (e.g. *flit* → *flitted*). Furthermore, connectionist models that learn purely *intra*-level phonological mappings cannot distinguish verb-stem

[2]See Plunkett & Marchman [1991] for details concerning the phonological conditioning of the suffix by the stem of the verb.

homophones that take different past tense forms (Pinker & Prince [1988]). For example, *to brake* and *to break* take past tense forms *braked* and *broke* respectively. Since the inputs to a phonological-phonological network in these cases are identical, so will their outputs remain identical. Homophones are unproblematic, however, when the acquisition of verb morphology is conceived as proceeding from semantics to phonology, as they constitute a many-to-one mapping.

However, two new potential problems arise. First, as (Pinker & Prince [1988]) point out, the semantics of a verb is *not* a good predictor of the type of inflectional mapping that it must undergo. The three verbs *hit, strike* and *slap* are closely related semantically but they have different mapping types relating their stem and past tense forms (*hit* → *hit, strike* → *struck, slap* → *slapped*). The network must learn to *ignore* this semantic similarity in learning the mapping. Second, this same problem arises in general for the network, insofar as there is an arbitrary relationship between the meaning of the item and its phonological form. Similar inputs do *not* lead to similar outputs. In particular, highly similar inputs, modeling synonyms, provide a potential source of error in these networks.

The problem of similar inputs is perhaps uninteresting from the point of view of a neural network researcher: After all, one simply needs hidden units to separate inputs that are similar. Our interest in this problem stems from the observation that this is essentially the problem posed to the child seeking to acquire adult forms of words: The similarity structures at the input and target level have no relation to one another (except in the case of morphological variants of the same word)[3]. This particular issue has not been studied, to our knowledge, in the connectionist literature, which tends to emphasize a different set of problems, such as how to learn parity more effectively. It is the cognitive modeling perspective that has led us to consider this kind of training set. We are thus less interested

[3]Of course, there is most likely some amelioration of this arbitrary relation as evidenced by the existence of sound symbolism. We ignore this here.

in whether the network achieves 100% effectiveness on the training set as we are in the stages the network passes through during learning. The former can depend a great deal on the particular learning method, parameters, and network topology. We suppose that the latter depends less on network learning than on the characteristics of the training set itself. That is, the data usually overwhelms the priors of the model. In neural network error-correction learning, it is often found that the network will attack the dimension in the data of highest variability first. Thus the network's stages of performance are driven by the data. In this work we explore the effects of the distribution of environmental samples on the surface forms produced by the network over the learning phase. That is, we are interested in the *developmental predictions* of our model.

Our goals in this work are:

1. To verify that the mapping from structured meanings to phonological forms can be learned by a recurrent network.

2. To explore the effects of input and target similarity on the acquisition of the forms.

3. To explore the ability of the network to learn the indirect mapping between the different surface forms of the same verbs.

4. To analyze the errors of the network, compare these to children's errors, and to make predictions about these errors in the light of the semantic representations used.

We report on two sets of simulations that differ in the nature of the semantic representations used to encode verb meanings. In the first case, we simply map from localist semantics to surface form in order to test the recurrent network's ability to learn the mapping. In the second case, we introduce a similarity structure on the input semantics that is arbitrarily related to the phonological similarity structure of the target patterns. In each case, we provide an evaluation of the performance of the network on trained verbs and of the ability of the network to generalize to verb forms on which it has not been trained. In the second set of simulations, we provide a detailed analysis of the effects of output and input similarity on the mappings acquired.

Before passing on to the simulations, however, we first review some of the empirical issues relevant to this modeling effort, and secondly, make explicit the assumptions that we believe are important (and those that are not important) in developing this model.

2 Issues

2.1 How does semantics influence the acquisition of adult forms?

The problem we are most interested in here is how the structure of the meanings influences the acquisition of adult forms. We should be clear here that we are not, in the current study, looking at how semantics is acquired. Rather, we assume that all of the relevant conceptual distinctions are already available to the learner (but see the discussion below on assumptions). Given this, we are then interested in how a structured representation of meaning might influence the surface forms produced.

The study that comes closest to providing data on this subject is a cross-linguistic study of the acquisition of locative expressions (Johnston & Slobin [1979]). Johnston & Slobin [1979] compared the performance of children between the ages of 2;0 and 4;8 in producing locative expressions (*in, on, under, beside, between, back* and *front*), in English, Italian, Serbo-Croatian and Turkish. While the main intent of their study was to evaluate the order in which these locative expressions were acquired in terms of their conceptual difficulty, it turns out that one interesting result of their study was that the more ways that a language has to express a particular relationship, the later children learning that language mastered the forms, relative to children learning a language with fewer means

Turkish	**Italian**	**English**	**Serbo-Croat**
8	11	16	25

Table 1: Number of different vocabulary items (types) used for the seven concepts *in, on, under, behind, in front, beside,* and *between* in four languages. Computed from (Johnston & Slobin [1979]), Table 2.

for expressing the same concept. Children were scored for expressing a locative notion if they used it more often in appropriate circumstances than inappropriate ones. One of their main findings was: "Children learning Turkish or Italian produced more different locative notions than those learning English or Serbo-Croatian." Compare this with the number of ways of expressing locative notions in the four languages (see Table 1).

In precisely the languages that had more ways to express locative notions, the children actually learned to produce *fewer* locative expressions when matched for age level. This is not so puzzling if we consider that in order to express essentially the same notions, Serbo-Croat children had to learn an average of 3.6 words compared to the Turkish children's 1.1 words. This means there are essentially 3.6 synonyms for each concept (*in, on, under,* etc.) in Serbo-Croat. Thus, the size of the vocabulary for each concept influences the pace of development. Here we see a direct influence of conceptual similarity on semantic differentiation and linguistic form. That is, synonyms are hard to learn.

2.2 What is the source of children's production errors?

We suggest that there are three general ways in which the current model makes predictions for children's production errors. The first concerns the basic architecture of the model and how that impinges on our predictions about error sources. A second is how error-correction learning leads to variations in production. A third is how the target vocabulary leads to predictions about syllabic structure.

One remarkable fact about children's language production is that they fail to produce some sounds that they had no difficulty producing during the babbling stage (Jespersen [1925]; Jusczk [1992]). That is, they have a rather full phonemic inventory during babbling, but when word production starts, a very different picture emerges. They start with a very small phonemic inventory, and slowly add new sounds. One possible account for this gradual accumulation of phonemes is that they may have the targets for word production in advance, perhaps even in the form of the motor program for the phonemes, but are having trouble associating the motor program with the meaning of the word. To quote Jespersen [1925] (taken from Dale [1972]):

> It is strange that among an infant's sounds one can often detect sounds – for instance k, g, h and uvular r – which the child will find difficulty in producing afterwards when they occur in real words ... The explanation lies probably in the difference between doing a thing in play or without a plan – when it's immaterial which movement (sound) is made – and doing the same thing in fixed intention when this sound, and this sound only, is required...

That is, they are "without a plan" for the word. This account naturally fits with the architecture of our model, which learns plans for the words, and assumes targets are available in advance. More specifically, on this view, the children lack a mapping from the plan specification to the sequence of phoneme motor programs.

One source for errors, then, will be in how hard this mapping is to learn. One can expect in advance, for example, that synonyms, which are essentially a one-to-many mapping for an architecture that maps from meaning to sound, will be difficult to learn. The model suggests a quite different account of the phenomenon of *overextension* to those typically suggested in the literature (Barrett [1978]; Clark [1973]; Plunkett & Sinha [1992]). Many children have been observed describing, for example, all four-legged things as "doggie".

This is often accounted for in terms of the semantic features associated with a given label. Underspecification of the list of semantic features associated with a label will result in that label being overapplied relative to the adult definition. As more features are acquired, more distinctions are made between the labels children apply to objects.

An alternative source of this type of error are the errors produced while learning the mapping between semantics and motor programs. Early in learning, the different features defining the children's early semantic representations are linked to different articulatory motor plans with varying connection strengths. For example, the feature FOUR-LEGGED associated with the semantic representations of CAT and DOG may be most strongly attached to the motor plan for producing the word "doggie". If other features that make up the semantic representation for CAT (e.g. MEOW) are not strongly attached to a different articulatory motor plan, then FOUR-LEGGED may win the competition for the phonological form used to express the particular meaning. Overextensions then result from the inappropriate outcome of competition between the different motor plans that are connected to any given semantic representation. Overextension is then viewed as arising from articulatory confusions as opposed to semantic confusions. Overextensions are eliminated as bundles of semantic features redistribute their patterns of connectivity to articulatory plans.

The second source of error is error-correction learning. In a network, error-correction learning leads to the discovery of the primary dimensions of variation in the training set before the detections of the lower level (fine-grained) distinctions. In this sense, the learning algorithm employed in the network performs an implicit principal components analysis. Thus, we expect the network to learn first the distinction between vowels and consonants since they are quite distant in the phonological representation used for the words in our simulations (see Figure 4). Individual phonemes will be discovered gradually by the learning algorithm. However, due to the fact that different phonemes must be mapped by the same set of weights, discovery of some phonemes will lead to renewed errors in others. Thus, we will see U-shaped learning as has been observed in recent simulation experiments (Plunkett & Marchman [1991]).

Stemberger [1990] has proposed that children's speech errors can be viewed as resulting from the operation of an error-correction system that is attempting to adapt to a perfect target. Errors result from simply trying to reproduce these targets. His model differs from ours, however, in assuming that between the semantic and pragmatic levels and the ultimate motor programming level, there are distinct lexical, syntactic, syllabic and phonemic levels. Our proposal suggests a single program level (the recurrent network) between a meaning level and the motor programming level. That is, the meaning level specifies the program for the network in generating the sequence of phonetic features which taken together make up the phonemes of the word. These are then presumably passed on to an articulatory level as specifications for programs there. Although we state our case rather strongly here, clearly, more than one level may be required for particular inflectional forms such as reduplication (Corina [1991]; Gasser [1994]).

We discuss the third source of error, the target vocabulary itself, in the next section.

2.3 The role of phonological form in molding children's vocabularies

In a case study of language acquisition in two Danish children between the ages of 12 and 26 months, Plunkett [1993] describes how children consistently use non-standard (idiosyncratic non-adult) forms to refer to objects and actions in the world. In some cases, these idiosyncratic forms could be related to target words in the language. Plunkett suggests that the children are having difficulty segmenting adult speech into words, and are *over-*

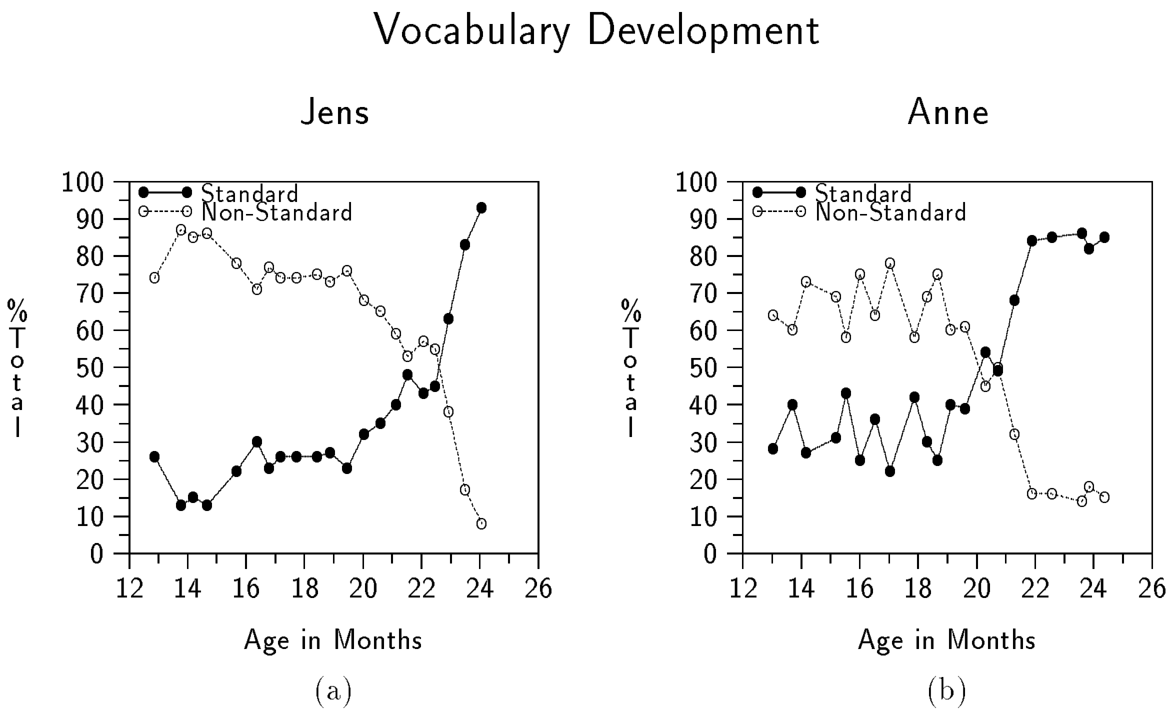


Figure 3: Vocabulary development in two Danish children: The number (types) of standard (adult-like) and non-standard (idiosyncratic) forms are plotted as a percentage of total vocabulary at different points in development during the children's second year. Based on data from (Plunkett [1993]).

shooting and *undershooting* the word boundaries. This explains at least some of his data. Other idiosyncratic expressions were unrelated to adult targets but were nevertheless used by the children in a meaningful and consistent fashion. Plunkett found an inverse relationship between the rate of use of these non-standard forms and adult forms in the child's vocabulary. Initially, the child's vocabulary was dominated by these non-standard forms, which then decreased as more adult forms were produced. This relation is depicted for the two children in Figure 3.

What influences the forms these pseudowords take? We have suggested above that one of the influences is semantics — similar inputs will initially lead to similar outputs. However, we also believe there are effects arising from the distribution of forms in the target language. Often error-correction learning algorithms find the average of the output first. This average will be weighted by the frequency of different forms. The network must then correct this overshoot to produce the less frequent forms. In our simulations, there are three syllable structures taken by stems in the target language. We will see below that this leads to predictions about the order of acquisition of syllabic structures which go beyond this simple simulation.

3 Assumptions

The essential assumptions that the model incorporates are:

1. The child must associate meanings and forms.

2. The child's representation of meanings and phonological forms each have their own intra-level similarity structure.

3. The similarity structures that characterise meaning and phonology are uncorrelated. I.e., similarity in meaning of two forms does not predict similarity at the phonological level of representation.

4. The child learns this mapping in an associative network.

5. Meaning includes features that correspond to morphological variants at the phonological level, such as tense.

The following assumptions we view as convenient and not essential to our conclusions:

6. The child has available perfect versions of the input and target forms.

7. The model uses back propagation learning with a sum-squared error criterion.

8. The model is learning an inflectional system that is analogous to the past tense system of English verbs.

The central concern of the model, therefore, is to evaluate the view that an important factor in determining the mapping from meaning to form is the influence of the similarity structures at the different levels of representation involved in the mapping. However, to avoid misunderstandings from the outset, we will discuss the last three assumptions that we have deemed inessential.

The child has available perfect versions of the input and target forms. That is, we assume the child has already divided his world up into categories and can distinguish them, and that the child has perceived the target forms correctly. Stemberger [1990] has argued that one can assume the child has perfect targets, and that most non-systematic errors can be explained in terms of a connectionist-style error correction learning rule. Even though we make this assumption here for convenience, we would not want to assume in general that the process of perceptual and conceptual organization has advanced to this point before the child's attempts to communicate commence. For example, Plunkett [1993] has argued that segmentation of the adult input into lexical items is not perfect and is certainly a possible source of error in children's forms. In particular, he hypothesizes that different children may choose to segment the input into smaller or larger chunks, and that this is reflected in their production vocabulary. Likewise, Clark [1973] has argued that the child's semantic representations develop in concert with vocabulary growth. In this paper, we will identify another possible source of error that is in alignment with Stemberger's proposal, but we do not believe it is the only source of error.

A second potential problem that follows from this assumption is that the model has available to it, from the beginning, a past tense marker in its semantic representation. Marcus et al. [1992] have hypothesized that the recognition of the obligatory past tense marker plays a crucial role in overregularization errors. While their position is controversial, we must agree that in our model, the marker is available at an overly early stage. However, it will only play a role in terms of morphological variations being acquired along with stem forms. This will not affect the major issue we wish to address, *viz.*, the issue of the effects of input and target similarity on acquisition. Again, we do not see this as an essential feature of our model, and could relax this assumption in future work. In fact, one could extend the model here to allow the inputs and targets to be learned even as the mapping is learned, and this would perhaps be more realistic. We prefer to deal with one issue at a time, however, and investigate the effects of learning the mapping without worrying for the moment about acquisition of semantics and phonology.

The model uses back propagation learning with a sum squared error criterion. We are not wedded to back propagation. It is simply a convenient method for training. Our choice of learning algorithm does have an effect on what is learned, however. In particular, under a rather loose set of criteria, any least squares estimator will learn the expected value of the output given the input (Duda & Hart [1973]). What this means is that initially, for very similar inputs, the network will produce averages of the target outputs. In our case, this will result in blends at the phonological level. Stemberger has observed semantic blends produced by children and adults, but for the most part, they are blends that start with segments from one word and then finish with segments from a related word, rather than blends on a phoneme-by-phoneme basis. However, there are systematic phonological errors that children make, and it is possible that some of these systematic errors have such blending as their source.

The model is learning the past tense. This is perhaps, the most simultaneously convenient and unfortunate assumption of our model: Convenient because it is a familiar domain that we have worked with before, and is certainly familiar to many readers by now: Unfortunate because it belies the generality of the model. This model might in theory be capable of representing the entire lexicon in the same network, although perhaps more plausibly with an appropriate mixture of experts architecture (Jacobs et al. [1991]) (see (Corina [1991]; Gasser [1994]) for alternatives).

4 Methodology

All simulations utilize a simple recurrent network of the type developed by Elman [1990] (see Figure 1). In all simulations the output phoneme consists of a 15 bit vector that reflects standard phonological contrasts. A noteworthy characteristic of this phonemic representation lies in its attempt capture the sonority relationships between vowels and consonants (see features O1–7 in Table 2).[4] A more perspicuous look at this representation is given in the cluster diagram shown in Figure 4. The task of the network is to output a sequence of phonemes that correspond to the stem or past tense of the verb whose semantic representation is presented at the input. This task is a bit of a historical accident: We view the network as a model of how *any* morphophonological regularities may be acquired. That is, with sufficient variation in the input semantics, different noun and verb forms could presumably all be produced by the same network. Thus, this is a potential model of the morphophonological processes for the entire lexicon, not just one variation in verb morphology.

The distinction between stem and past tense forms is encoded by a 2-bit vector at the input level. The inventory of verbs (both stems and past tense forms) that the network is required to produce at the output is taken directly from previous simulations conducted by (Plunkett & Marchman [1991]), though note that a different phonological representation is used. In the current work 504 stem/past tense pairings are used. Each stem consists of a Consonant-Vowel-Consonant (CVC) string, a CCV string or a VCC string. Each string

[4]The phonological representation presented here was originally designed by Alan Prince and Kim Plunkett.

				LO	MD	HI	GL	SN	FR	ST	PLACE					
		BK	TN	O7	O6	O5	O4	O3	O2	O1	LA	CR	VL	NS	SB	VO
vat	æ	−1	−1	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
cut	ˆ	+1	−1	+1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
ate	e	−1	+1	−1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
bet	ε	−1	−1	−1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
boat	o	+1	+1	−1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
bought	O	+1	−1	−1	+1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
beat	i	−1	+1	−1	−1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
bit	I	−1	−1	−1	−1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
boot	u	+1	+1	−1	−1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
foot	$\mho$	+1	−1	−1	−1	+1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
	y	−1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
	w	+1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1	−1	+1
	h	−1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	−1	−1	−1	−1
	r	−1	−1	−1	−1	−1	+1	+1	+1	+1	−1	+1	−1	−1	−1	+1
	l	−1	−1	−1	−1	−1	−1	+1	+1	+1	−1	+1	−1	−1	−1	+1
	m	−1	−1	−1	−1	−1	−1	+1	+1	+1	+1	−1	−1	+1	−1	+1
	n	−1	−1	−1	−1	−1	−1	+1	+1	+1	−1	+1	−1	+1	−1	+1
ng in ring	η	−1	−1	−1	−1	−1	−1	+1	+1	+1	−1	−1	+1	+1	−1	+1
	f	−1	−1	−1	−1	−1	−1	−1	+1	+1	+1	−1	−1	−1	−1	−1
	v	−1	−1	−1	−1	−1	−1	−1	+1	+1	+1	−1	−1	−1	−1	+1
	s	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	+1	−1	−1	+1	−1
	z	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	+1	−1	−1	+1	+1
theatre	θ	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	+1	−1	−1	−1	−1
mother	δ	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	+1	−1	−1	−1	+1
	p	−1	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	−1	−1	−1	−1
	b	−1	−1	−1	−1	−1	−1	−1	−1	+1	+1	−1	−1	−1	−1	+1
	t	−1	−1	−1	−1	−1	−1	−1	−1	+1	−1	+1	−1	−1	−1	−1
	d	−1	−1	−1	−1	−1	−1	−1	−1	+1	−1	+1	−1	−1	−1	+1
	k	−1	−1	−1	−1	−1	−1	−1	−1	+1	−1	−1	+1	−1	−1	−1
	g	−1	−1	−1	−1	−1	−1	−1	−1	+1	−1	−1	+1	−1	−1	+1
silence	-	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

Table 2: Phonological representation: The coding scheme is to be interpreted as follows: LO = low, MD = medium, HI = high, GL = glide, SN = sonorant, FR = fricative, ST = stop, BK = back, TN = tense, O7–O1 = sonority, LA = labial, CR = coronal, VL = velar, NS = nasal, SB = sibilant, VO = voiced.

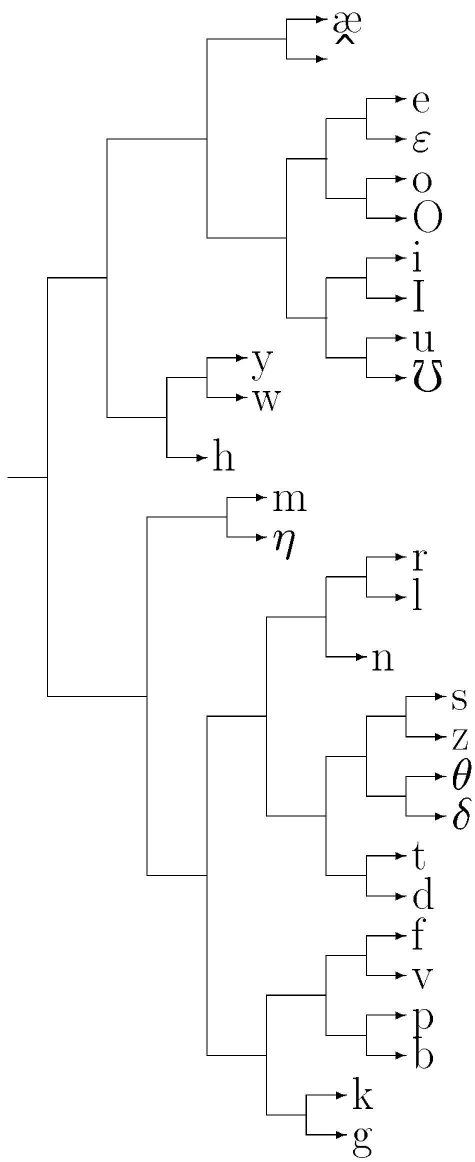


Figure 4: A Cluster Analysis of the Phoneme Representations: Note that vowels are clustered in a group separate from the consonants and that the semi-vowels y and w inhabit an intermediate position in the clustering hierarchy. Consonants are paired up according to their voicing contrasts. n seems to be placed inappropriately far from m and η. However n shares the feature +coronal with r and l.

is phonologically well-formed, even though it may not correspond to an actual English word. Verbs are assigned to one of four classes. Each class corresponds to a different type of transformation analogous to a distinct past tense form in English as in (Plunkett & Marchman [1991]). The first class follows the Regular rule in English. The three irregular classes are: (1) *Arbitrary verbs* have no systematic relationship between the stem and past tense form (*go* $\rightarrow$ *went*), (2) *Identity verbs* are unchanged between forms (*hit* $\rightarrow$ *hit*), and (3) *Vowel Change verbs* undergo a change in vowel in CVC forms (*strike* $\rightarrow$ *struck*). There were 2, 20 and 20 verbs in each of these classes, respectively, with frequecies of 5, 1 and 2 times that of the regular verbs. Verbs are assigned randomly to each of the four classes, with the constraint that stems possess the appropriate characteristics of a given class.

Semantic representations of verbs are of two types. In the first set of simulations, each verb is represented in a localist manner in a 500-bit vector. An additional two units encode whether the network is to produce the stem of the verb or the appropriate past tense. In the second set of simulations a similarity structure is imposed on the semantic representations by using distortions of several prototype vectors (Chauvin [1988]; Knapp & Anderson [1984]). Distortions may vary in their distance from the prototype. We use 50-70 prototype vectors (and thus as many categories), with an appropriate number of distortions to provide an input representation for every word. Note that each distortion corresponds to an individual word meaning, rather than as a "noisy" example of a single word's meaning. Thus similar inputs will have very different outputs.

The amount of distortion may vary. Thus we can essentially build a "semantic network" of concepts, which corresponds to the structure of the semantic space that would be induced by a recursive clustering technique. In this paper, we explore an extremely simple semantic network structure. We only do one step of distortion of a prototype, corresponding to a very "flat" semantic network. The amount of distortion around each prototype is held constant, but varies across prototypes to produce high, medium, and low distortion clusters. The low distortion clusters we term the *synonyms*.

Training proceeds by randomly selecting a plan vector (the verb's semantic representation) and a tense bit (stem or past tense). This composite vector is then presented at the input units over a number of time steps that correspond to the number of phonemes in the output form. All stems consist of three phonemes and thus involve three time steps. Past tense forms may involve 3, 4 or 5 phonemes. At each time step, the discrepancy between the actual output of the network and the target phoneme is used as the error signal to a back propagation learning algorithm. We use a version of the TLEARN simulator developed by Jeff Elman at UCSD, modified to run on the Cray Y-MP. As part of the teaching signal, the verb plan is trained to produce an end-of-form signal (corresponding to the *silence* phoneme in Table 2). The "context units" are reset between forms. We used a learning rate of 0.1, and no momentum.

5 Analysis

The performance of the network is analysed at regular intervals in training. In this paper we present two types of analysis. First, we determine the hit rate for stems and past tense forms, both on the entire training set and on a class-by-class basis. The phoneme produced at any time step is determined by simply taking the closest legal phoneme to the output in terms of euclidean distance. This is similar to the settling phase of an attractor network (cf. (Plaut & Shallice [1993])). An output must match the target on all phonemes to be counted as a hit. In the error graphs we report whether the output sequence is a hit or a miss. In the second set of experiments, we carry out further analysis of the outputs actually produced.

We analyse the generalization characteristics of the network by first training the network with 27 verb plans to produce only the stem form of the verb and with another 27 verb plans to produce only the past tense form of the verb. Each verb plan is then tested on the phonological form of the verb to which it has not been trained, e.g., 27 stem forms and 27 past tense forms. The output of the network on these novel inputs is used to evaluate the network's generalization properties.

Finally, in the second set of simulations we provide a detailed analysis of the output of the network and relate the role of semantic similarity to the similarity of the phoneme sequences across different verbs, and the syllabic structure that the network extracts over the sequence of output phonemes.

6 Experiment One

This experiment reports the results of simulations using a 500 word vocabulary and orthogonal representations of the verb plan. Figure 5 (a) provides a summary of the network performance on all past tense forms and all stem forms while Figure 5 (b) compares the generalization characteristics of the network in predicting the past tense forms of the verb when it has only been trained on the stem and *vice versa*. Figure 5 (a) shows that the network is equally fast at learning both stem and past tense forms and that learning undergoes a spurt in growth around the 20 epoch mark. In contrast, the test verbs differ with respect to their performance on stems and past tense forms. Figure 5 (b) shows that when a verb plan is trained to a past tense form, the network is quite accurate in predicting the correct stem ($\geq 90\%$ after 70 epochs of training). On the other hand, generalization from stem to past tense never exceeds 55% over this training interval. It should be noted that we use a very strict criterion for generalization: All past tenses are assumed to be Regular. Over several simulations starting with different random seeds, we find that performance on past to stem generalizations is always good, while stem to past varies. This result is

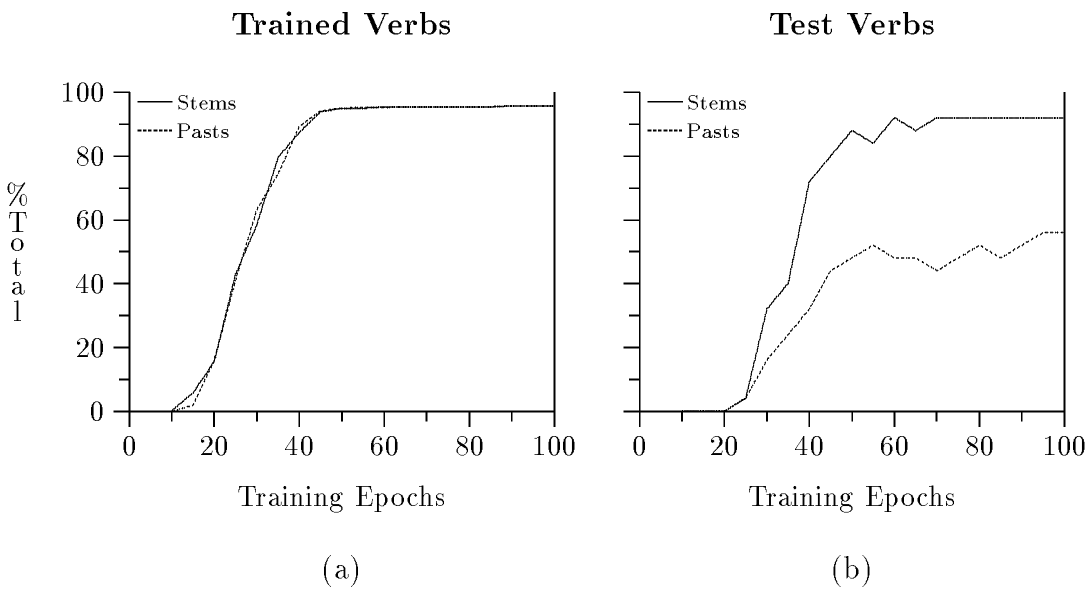


Figure 5: Overall results for experiment one (orthogonal semantic representations). Panel (a) shows the performance on the training set, for verbs trained on stem or past. Panel (b) shows generalization performance, which is much better for verbs trained on only the past form than for verbs trained only on the stem form.

to be expected given that the form of the past tense is a better predictor of the stem than the stem is of the past tense form (e.g. if the past tense of a verb is *talked* then its stem form is unambiguous, but if the stem is *hit* then, in principle, its past tense form is underdetermined). Indeed the discrepancy between the generalization curves in Figure 5 (b) can be accounted for in this fashion (see Figure 7 (c) below).

Figure 6 (a) provides a class-by-class breakdown of network performance on trained stems. These result indicate that the Regular, Identity mapping and Vowel Change classes are learned first, while Arbitrary mappings are delayed. Figure 6 (b) reveals a similar rank ordering of classes with past tense forms.

Figure 7 provides a breakdown of the major error types. The lack of errors early in training in these graphs is due to the fact that "garbage" outputs are not captured by any of the following classes of errors. The graphs show that the network has some difficulty with the phonologically-conditioned ending on the Regular class. However, many of these

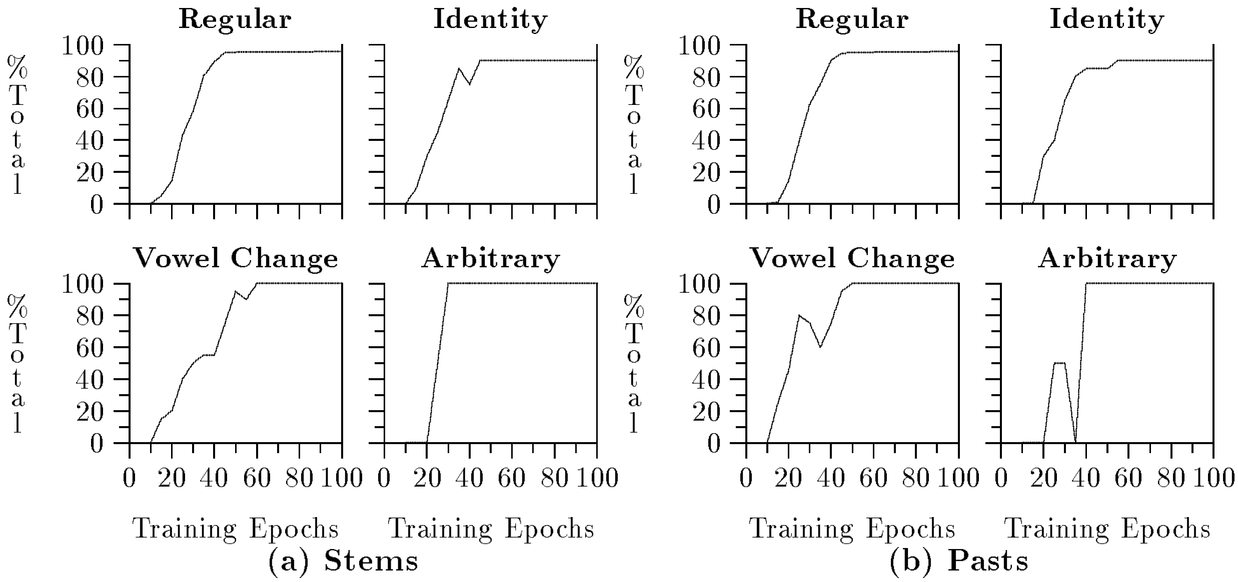


Figure 6: Percent correct on training set, by form class (orthogonal simulations). Panel (a) shows the performance on trained stems, panel (b) shows performance on trained past forms.

errors correspond to difficulties in producing the epenthesized form of the Regular past. The network apparently never overregularizes an irregular form, although it will produce the wrong type of irregular mapping. More analysis of this point is given in the next Experiment. Finally, we classified an error as "Phonological" if one of the phonemes of the word was within one phonetic feature of the correct output. Thus these are near misses, and are produced as the vocabulary burst is just taking off.

7 Experiment Two

This experiment reports the results of simulations using a 504 word vocabulary and semantically structured representations of the verb plan. This poses a much more difficult problem for the network, and a more interesting one from a psychological viewpoint. Since our main interest in this section is on the effects of input and output similarity on acquisition, we begin with an analysis of target similarity effects on the vocabulary profile,

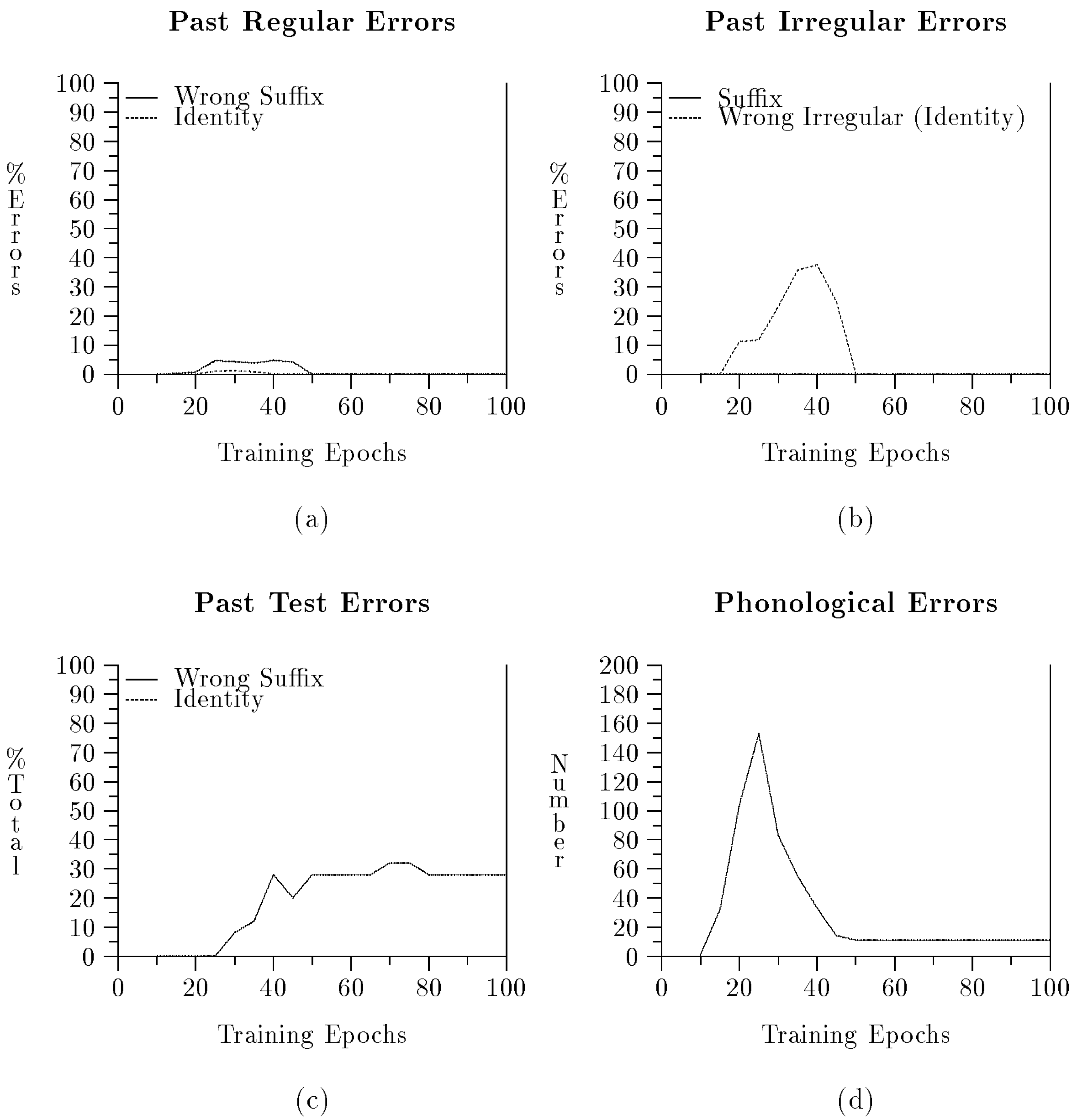


Figure 7: Error forms by class (orthogonal simulations). Panel (a) shows errors on Regular pasts for the training set. Panel (b) shows errors on irregular pasts for the training set. These turned out to be incorrect Identity mappings. Panel (c) shows generalization errors on Regular pasts that were trained on Regular stems only. Panel (d) shows Phonological errors (see text) on the trained verbs.

then turn to the effects of semantic structure. Finally we analyze the performance of the network, with an emphasis on how well the indirect mapping from stem to past is learned. We thus perform four kinds of analyses on this network:

1. We measure the changes in the stem output strings during learning from the point of view of the syllabic structure of the target language.

2. We measure the changes in similarity of stem output strings during learning with respect to the semantic clusters.

3. We measure the simple performance of the network in terms of hit rates and error rates.

4. We measure the rate at which the indirect mapping from stem to past is learned irrespective of the correctness of the stem form.

7.1 Syllabic structure changes

The syllabic structure of the stem forms in the artificial language consist of CVC, CCV, and VCC forms. In rough accord with their distribution in English, there are 368 CVC's, 63 VCC's, and 46 CCV's.[5] In order to assess the vocabulary development of the network, we divide the stem output strings of the network into three classes:

Words: Strings that belong to the target vocabulary.

Pseudowords: Strings that are not in the target vocabulary but conform to the syllabic structure of the language — CVC, VCC, or CCV.

Non-words: Strings that do not fit the above criteria — CCC, CVV, VCV, VVC and VVV.

[5]These numbers only add to 477 because 27 stems are left out of the training set for testing purposes.

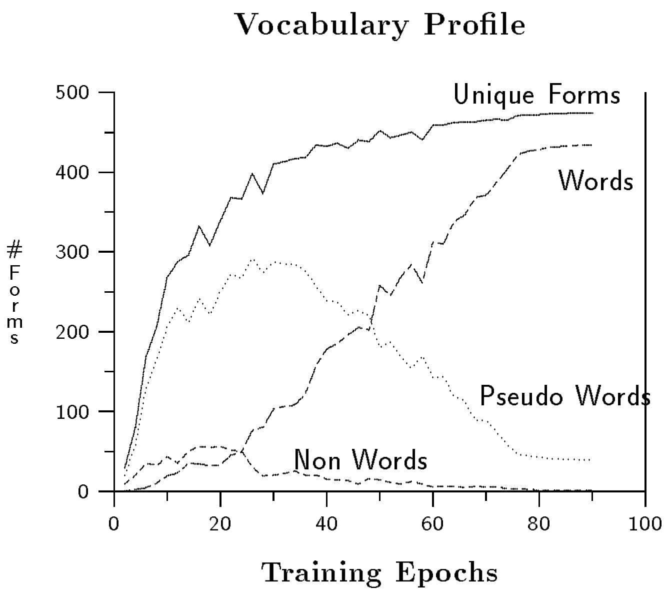


Figure 8: Vocabulary profile of the network "language" (compiled from the training set stem outputs) over training. "Unique Forms" is just the sum of the other three curves. See text for definitions of the classes.

A graph of the numbers of unique forms of each kind over learning, along with the total number of unique forms, is shown in Figure 8. Interestingly, before it has acquired many words in the language, the network has captured the syllabic structure, as evidenced by the high proportion of pseudowords in the set of unique forms. That is, there is a "private" vocabulary burst before the burst in adult forms later. Over learning, there is an inverse relationship between these forms and the correct forms as the pseudowords migrate into the target vocabulary. This pair of curves is reminiscent of the set of curves found by Plunkett [1993] (see Figure 3). The total number of consistently produced non-(adult) Danish forms was inversely related to the number of Danish words over the period studied. That is, the children had their own vocabulary early in development that was eventually replaced by target forms.

Note that in the simulation, the number of nonwords actually *increases* just prior to the onset of the acquisition of adult forms (around 10–20 epochs). Why should this happen? Further analysis reveals a simple explanation for this effect. We assign each string of the

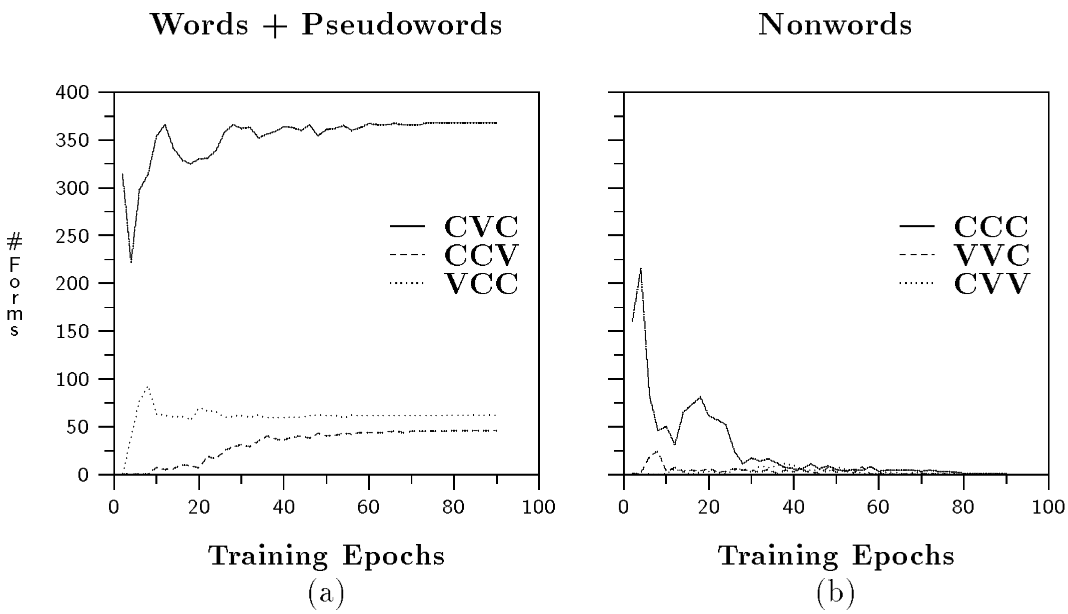


Figure 9: Syllabic structures of the network outputs on the training set over learning. Panel (a) shows the count of syllable structures that are correct for the language (irrespective of whether the outputs themselves match their targets). Panel (b) shows the number of incorrect forms.

network's stem outputs to one of the eight possible combinations of 3 segments of C's and V's regardless of whether the form is an adult form or not. Three of these classes characterize the target language's syllabic structure. A plot of the number of network outputs that belong to the three legal classes is given in Figure 9 (a). Notice that the number in each class over- and under-shoots the correct number several times during acquisition, but the largest swings are early in learning. In particular, no CCV's are produced early on, and too few (and then too many) VCC's. That is, the network quickly starts to produce the most dominant form in the language, CVC, then must restructure to obtain the correct proportions. In order to extend to the target language syllabic structure these strings have to mutate from CVC to CCV and VCC (and sometimes back again) during epochs 2 through 80. Suppose that the network only changes one segment at a time. Strings changing from CVC to CCV have to change the mid-vowel to a consonant and the coda consonant to a vowel. The possible intermediate forms are CCC and CVV.

Similarly between CVC to VCC, the possible intermediate forms are CCC and VVC. Indeed, a graph of the number of these forms produced by the network (Figure 9 (b)) shows that they occur only during the cross-over from pseudowords to words. Also the form that is on both paths, CCC, is the most common nonword produced.[6] The (logically possible) forms VVV and VCV never occur.

Thus these simulations make the relatively unintuitive prediction that children may produce forms that are outside the target language yet which conform to the syllabic structure of the target language. This prediction is consistent with the finding (Plunkett [1993]) that children produce non-standard pseudowords during early language acquisition. A further prediction is that the predominant syllabic structure of these early pseudowords will concur with the dominant syllabic structure of the language. The migration of these syllabically legal, but non-standard, forms to the correct target forms may result in a transitional period where illegal syllabic structures are generated. Those syllabic forms that are on more "paths" between adult forms should be produced more often. Of course, not all illegal syllabic forms will be produced by the child as they may be inarticulable (e.g., CCC). The current status of research on child phonology does not permit us to properly evaluate these predictions, though recent work in the area (Vihman et al. [1985]) tends to emphasize the constructive role of the child in the acquisition of its phonological system — a perspective consistent with current predictions.

7.2 Input Similarity Effects

The semantic structure consists of a number of semantic clusters, or groups. These are 100 element vectors formed in the following way: First, a prototype vector is formed by

[6]We assume that the output of the model is actually a specification of a motor plan for that word. Presumably the articulators would prevent the production of CCC's.

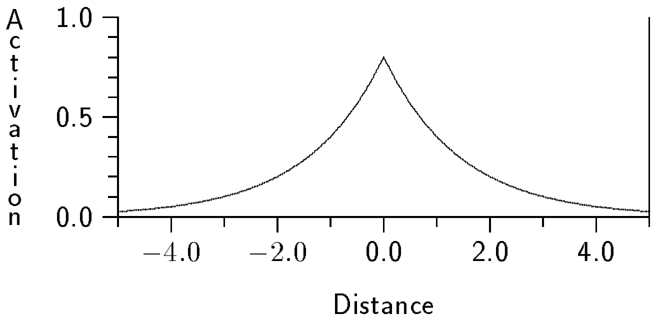


Figure 10: Pre-processing function for convolving distorted patterns.

randomly choosing a certain number of bits to be "on" in the prototype. Then, distortions of this prototype are generated by randomly shifting these bits in a uniform distribution with a specified range around its current position in the vector. A distortion level of "1" corresponds to randomly selecting with probability 1/3, the current position or one of the neighboring positions in the vector. Finally, the vectors are convolved with an exponentially decaying function shown in Figure 10 which enhances the similarity structure of the distortions within a prototype cluster. The whole three stage process starting with the prototype pattern through the distorted pattern to the convolved distorted patterns is exemplified in Figure 11.

The structure of each semantic cluster is varied along two dimensions. First, the number of 1's in the prototype is varied from 10–15. This is essentially varying the "abstractness" of the categories (Plaut & Shallice [1993]). Second, the amount of distortion is varied between categories, so some categories are high distortion (level 6), medium (level 4) or small (level 1). We operationally define the level 1 distortion categories as the *synonyms*.[7] In the following, we had 3 synonym classes, 45 medium distortion groups and 24 high distortion groups.

This semantic structure has the advantage of avoiding the problem of defining exactly what the features represent, but has the disadvantage of being an exceptionally arbitrary

[7]These might better be called near-synonyms.

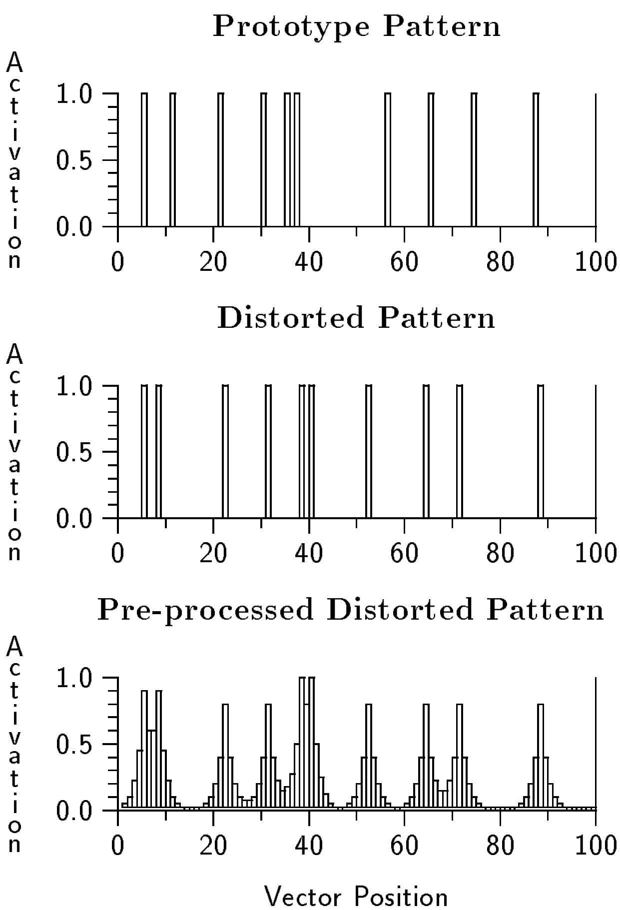


Figure 11: The creation of a meaning pattern. Top: A binary "prototype" vector is created. Middle: The position of the bits in the vector are jittered to create a distortion of the prototype. Bottom: The distorted pattern is convolved with the function in the previous figure to produce the final "meaning pattern". This process was based on the one used by Chauvin [1988].

and probably unrealistic structure. That is, it is unlikely that actual human categories have this structure, but for the moment, we are simply interested in what effects these variables have on our model. More realistic semantic structures, with hierarchies and non-uniform dispersion, are left for future work.

We hypothesized that the synonym groups would produce outputs that were more similar to one another than the non-synonym groups and the vocabulary as a whole. In order to test this hypothesis, we must have a quantitative measure of group similarity. If we use the absolute distance between the output vectors, this ignores the chance distribution of words (targets) to a group, which can result in more or less similar outputs. Instead, we consider the distance between output vectors *relative* to the distance between their targets:

$$\mathcal{RD}_{i,j} = \frac{dist^O_{i,j}}{dist^T_{i,j}}$$

where $dist_{i,j}$ is the sum of the distances between the corresponding phonemes of words i and j, using euclidean distance between the vector representations, and the superscripts denote target and output. This is the distance between the output phonemes in a string *relative* to the distance they should be after learning. $\mathcal{RD}_{i,j}$ should tend to 1 as the network learns, since as the outputs approach their target values, the distances between them will approach the distances between the target values. If the network outputs the same string for words i and j, $\mathcal{RD}_{i,j}$ will be 0. If it produces the correct string for each, $\mathcal{RD}_{i,j}$ will be 1.

For each group of interest, we average this measure across all the pairs in the group, and subtract it from 1 to obtain a similarity measure:

$$\text{Similarity}(\mathcal{G}) = 1 - \frac{1}{N(N-1)} \sum_{\substack{i,j \in \mathcal{G} \\ i \neq j}} \mathcal{RD}_{i,j}$$

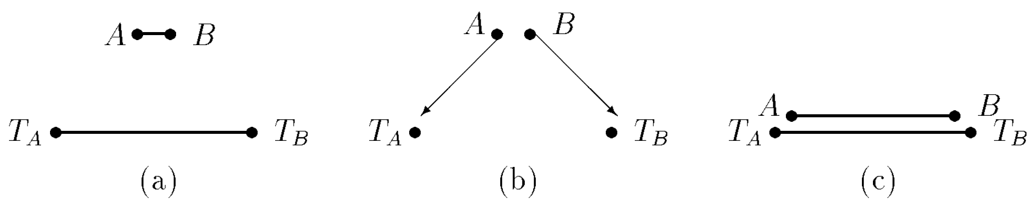


Figure 12: An illustration of the relative distance measure as learning proceeds. Initially, (a), A and B are the outputs from similar semantic vectors which have distinct phonological target T_A and T_B. The relative distance measure is the ratio $\frac{AB}{T_A T_B}$ which is small. In (b), the outputs have moved closer to their targets until at (c) the ratio is closer to 1.0.

where i, j range over the members of $\mathcal{G}$, and $\mathcal{G}$ is of size N. If the network produces the same output for every member of G, then Similarity(G) will be 1. As the words outputs approach their target values, Similarity($\mathcal{G}$) will tend to 0. This is illustrated graphically in Figure 12.

We apply this similarity measure to the strings produced by the network for each of the 72 prototype classes. Figure 13 (a) shows the average of this measure across the low-distortion (synonym) classes and the average across the high- and medium-distortion classes. Figure 13 (b) shows the similarity measures ($\mathcal{G}$) when the similarity score of the *total* output of the network has been subtracted from each group score. The curves show that, in general, semantic classes produce surface forms that are more cohesive than the target forms of the network are as a whole. As it should, this effect disappears over training. This is the result of the network overcoming the false cue of input similarity.

As expected, the synonym classes have higher within-group similarity throughout training than non-synonym classes. It is noteworthy that during the period that the network begins acquiring the target vocabulary, i.e. from 20–40 epochs (cf. Figure 8), the within-group similarity of the synonyms *increases* compared to the rest of the classes. That is, synonyms are forced to sound more alike than they did previously and more alike than the rest of the vocabulary at the beginning of acquisition. This is true on an absolute scale as

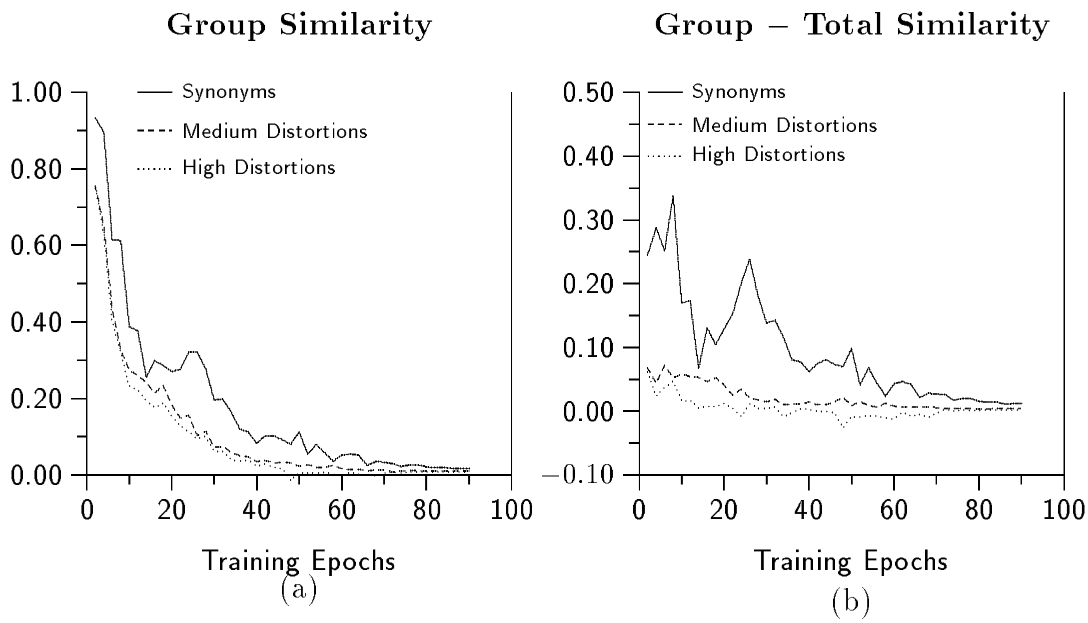


Figure 13: Input Similarity Effects. Panel (a) shows Similarity($\mathcal{G}$) plotted for the three semantic groups on the trained stem outputs (see text). Panel (b) shows the same data, with the normalizing factor of Similarity(Total Vocabulary) subtracted off. This emphasizes the relative increase in "Synonym" group similarity between 20 and 40 epochs.

well (where the relative similarity of the total vocabulary has not been subtracted off), as is shown in Figure 13 (a). This Figure also shows that even though there are many unique forms early on, as was shown in Figure 8, they are quite homogeneous. The explanation for the synonym effect in terms of network learning is simply that patterns that are more easily learned (because they have less input similarity) are dominating the error gradient. The patterns may be characterized as *competing for representational resources* at the plan (first hidden) layer. The more similar inputs are pushed together at this layer while the network carves out representations for the more disparate meanings.

Examination of the network outputs over the training period reveals that many of the output strings for synonym classes during the second peak of similarity at 24 epochs are within 2 or 3 features of one another. An interesting question here is: What is the string the outputs of a synonym class are pushed towards? Is it a blend of all of the strings of the class, or does one string in the class "capture" the output for that class? We tentatively find that when there is one lexical item that is more frequent than the others, it tends to capture the class. Interestingly, this is not the case if the more frequent item in a synonym class is an Arbitrary verb (*go/went*, probably due to the fact that the semantic representations of Arbitrary verbs map similar inputs to very different outputs (see Bartell, Cottrell & Elman [1991] for a thorough discussion of this issue). In the case that synonym verbs are equally frequent, a blend of all of the outputs for the class is produced. This counterintuitive, and possibly simply wrong, prediction of our model in large part depends on the use of a mean-squared error criterion, which tends to produce the average of outputs for the same input. A different learning rule, such as contrastive hebbian learning (Movellan [1990]) that learns distributions rather than expected values, could lead to a more satisfying result.

Taken in combination, these results suggest an unorthodox account of the source of the non-word forms found in Plunkett's subjects. These consistently used pseudowords are the result of two constraints or pressures on the child's language production: (1) A pressure to produce forms that are in keeping with the syllabic structure of the language at the output level, and (2) a pressure to produce similar forms based on input similarity. The child is thus producing the best approximation to a word in the language that is a blend of *all* of the words for that semantic class, with a tendency for this blend to be similar to the most frequent element of that class. A second counter-intuitive prediction of this work is that, during acquisition of the correct forms, the child will produce strings that may be inappropriate for the target language because they are *between* a common (over-acquired) form and a less common form.

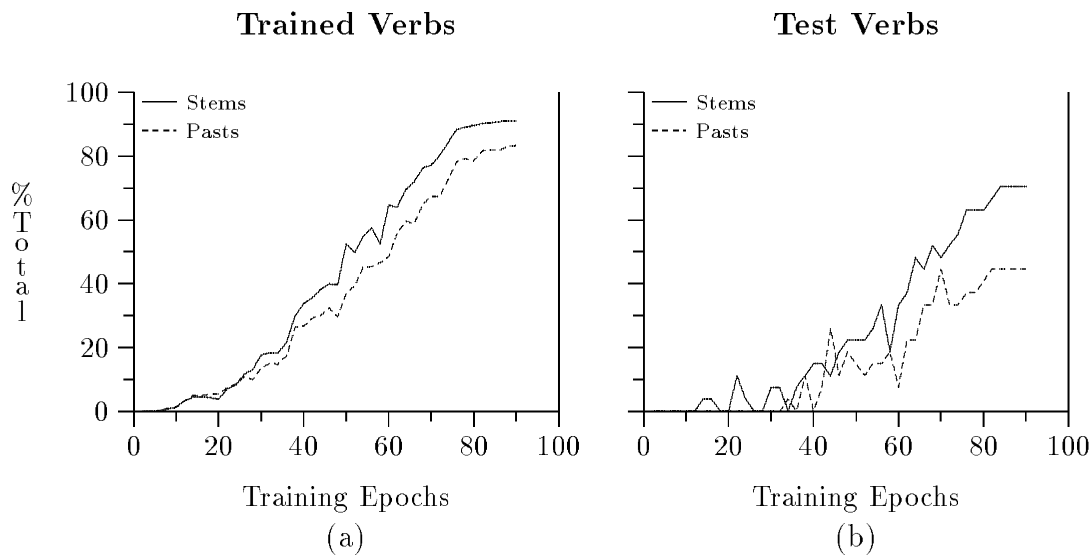
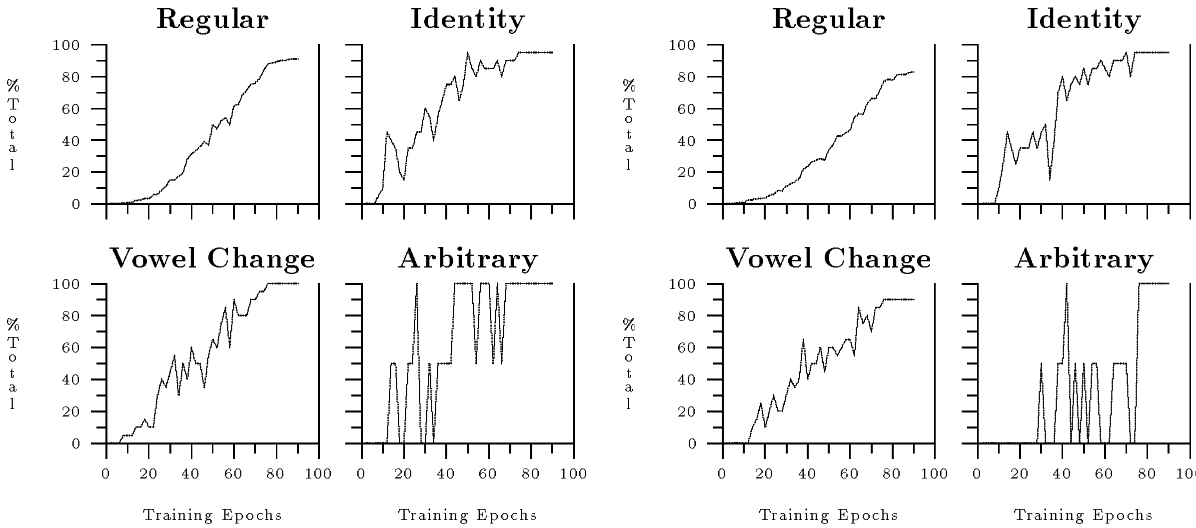


Figure 14: Percent correct from the semantic simulation. Panel (a) shows performance on the trained verbs. Panel (b) shows generalization performance on the untrained forms.

8 Performance Analysis

Figure 14 (a) provides a summary of the network performance on all past tense forms and all stem forms while Figure 14 (b) compares the generalization characteristics of the network in predicting the past tense forms of the verb when it has only been trained on the stem and *vice versa*. Unlike Experiment One, Figure 14 (a) shows that the network is slightly faster at learning the stem versus the past tense forms. Learning undergoes a spurt in growth around the 20 epoch mark. Just as in Experiment One, there is a contrast between the test stems and the test past tense forms. However, the generalization characteristics for this network are more restricted than in the simulations where an orthogonal verb plan is used. An examination of the output for test stems indicates that the network has difficulty generating the epenthesized form of the */ed/* suffix.

A breakdown of the performance of the network on the trained stem form of the different verb types is given in Figure 15 (a). Similar results are shown for the past forms, as seen in Figure 15 (b). The difference in the smoothness of the curves can be attributed to

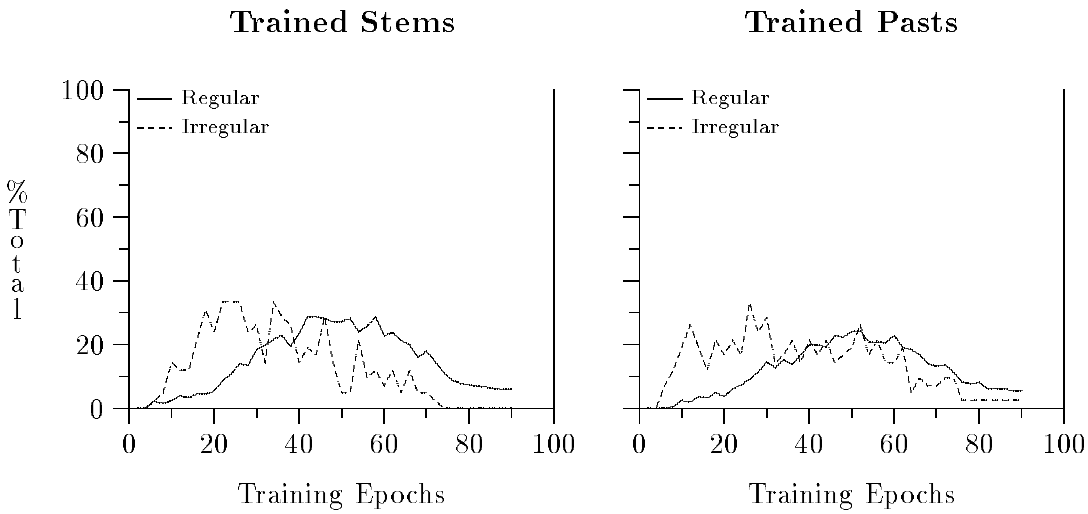


(a) Stems (b) Pasts

Figure 15: Percent correct on the training set by form class (semantic simulations). Panel (a) shows performance on the stems, (b) on the past forms.

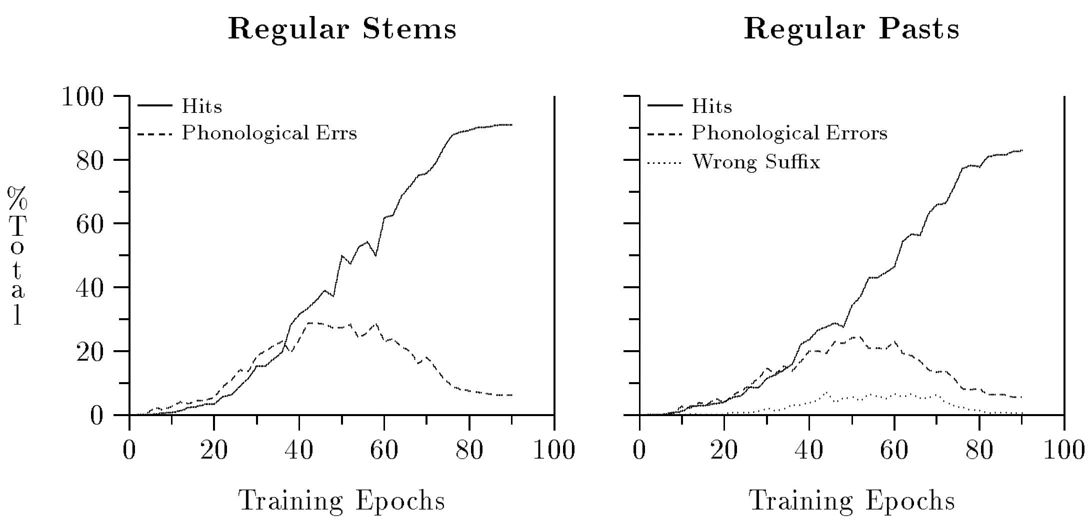
the size of each class, and the difficulty of the Arbitrary mapping. As a percent of class size, the network learns the Identity verbs the earliest, followed by the Vowel Change and Arbitraries. The advantage of the Identity class can be attributed to two factors: Their mapping is "supported" by the Regular mapping, where the stem portion is unchanged by the past tense transformation, and second, they are a simple mapping. In both the past and stem forms, the network simply has to produce the same output. Since the inputs are very similar in these two cases, it makes the mapping easy to learn.

Phonological errors in the semantic simulations are shown in Figure 16. Our criteria for phonological errors are so strict (only one phonetic feature incorrect on one phoneme) that this graph basically shows how many "near-misses" we have. This is illustrated in Figure 17, where we plot the hit rates along with the phonological errors for Regular verbs. We also show in Figure 17 (b) the one other classifiable error type for Regular pasts, adding the wrong suffix. This corresponds to adding a voiced final consonant in an unvoiced environment and vice-versa. Since there were so few Regulars ending with an



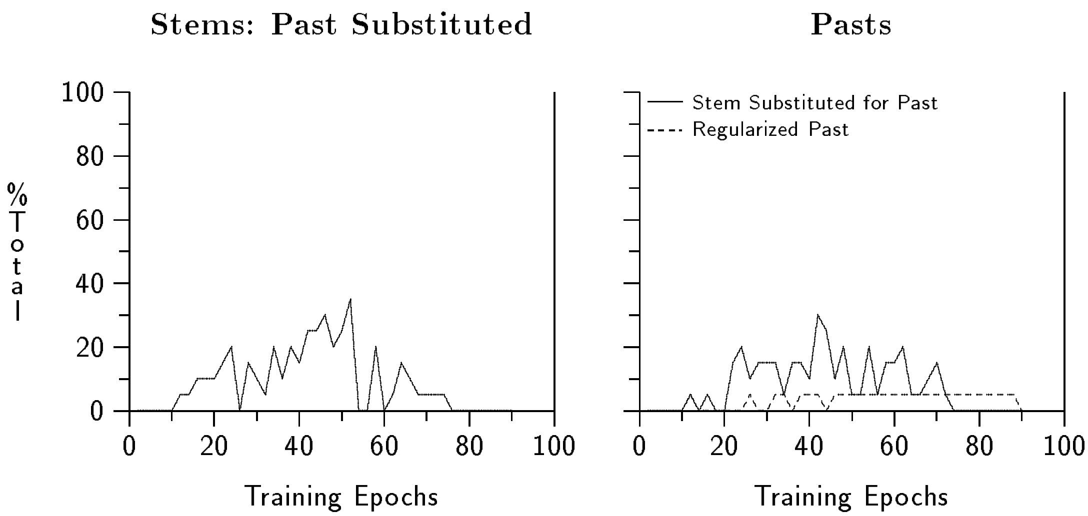
(a) (b)

Figure 16: Phonological errors in the semantic simulation on the training set.



(a) (b)

Figure 17: Phonological and suffix errors on trained Regulars. No Regular stems were given a suffix.

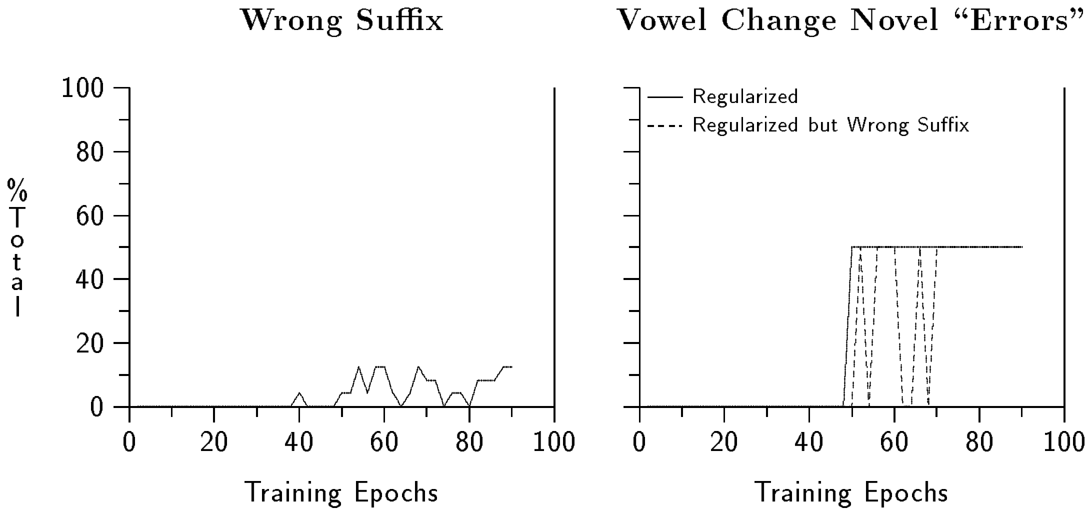


(a) (b)

Figure 18: Non-phonological errors on trained Vowel Change verbs. Panel (a) shows the number of stems for which the past form was substituted. Panel (b) shows the number of pasts for which the stem was substituted, and the number of (over)-regularized pasts.

epenthesized vowel, the network never overgeneralized this ending. Furthermore, it had a great deal of difficulty with these forms when appropriate.

The only classifiable errors for the irregulars aside from phonological ones were in the Vowel Change class. These are shown in Figure 18. The two major error types are a reverse-Identity mapping, shown in Figure 18 (a), where the past form was substituted for the stem form, and the converse, shown in Figure 18 (b), where the stem form was produced for the past. This reflects the difficulty of producing a mapping where the stem of the verb changes from one morphological form to another in an environment where 96% of the forms do not do this. The other error shown is a very unexpected one. Although in earlier simulations we had included homophones and found they were no problem for the network, we inadvertently included one in this simulation, in a Regular form and a Vowel Change form. The error shown corresponds to regularizing the past form of this word. Thus in this case we find that output similarity has resulted in similar plans for these two words, causing the past form to be regularized.



(a) (b)

Figure 19: Non-phonological errors on novel verbs. All are assumed to be Regular. Panel (a) shows that the voicing rule is not always followed. Panel (b) shows the results for two novel verbs that satisfied the phonological criteria for the Vowel Change class. Both were regularized, one with the incorrect Regular ending.

The phonological errors for the novel verbs followed the same pattern as for the trained verbs. In Figure 19 (a) we show that there is a small percentage of the Regular novels where the voicing rule was not followed. We also included two VC forms in the novel training set. They were both regularized, one with the correct ending and one with the incorrect ending (Figure 19 (b)). The incorrect ending was on a VC that ended in a dental, so that the (low frequency) epenthesized form was the correct answer.

9 Mapping Analysis

In this section we analyze how well the network acquires the indirect mapping described in Figure 2 as MAPPING 2. Although it may appear to the reader that we have already accomplished this in the previous section via a conventional "hit rate" analysis, we will soon see such "hit rate" analyses on this kind of network can be misleading. Second, we will show that the indirect mapping can be learned independently of the ability of the network to produce the correct forms.

The approach taken here is to analyze the output of the network in terms of a "hit rate" for past tense forms *relative* to stem forms, irrespective of whether the *actual* stem is correct or not. This is to be contrasted with scoring the past output based on the *target* specified for the stem. For example, if the network produces */w^g/* when the stem is called for, we will score a Regular hit if it produces */w^gd/* when the past is required. Since the past tense form is now evaluated relative to the network-generated stem form, we can classify the types of indirect MAPPING 2 relationships that the network has encoded.

Figure 20 shows a four-way classification of the network-defined indirect mapping — Regular (add a suffix), Vowel Change, Identity mapping and "other" (essentially a garbage class). These graphs show each kind of mapping as a percent of the total vocabulary. The proportion of Vowel Change and Identity mappings reflect the small number of such contingencies in the training set. The so-called "other" mapping corresponds to no classifiable relation between the two forms in this case. We see that the Regular mapping constitutes a substantial proportion of the MAPPING 2 possibilities and is actually acquired much earlier than many of the vocabulary words! This again reflects the fact that the model has available a reliable morphological marker from the beginning of learning. The low fraction of incorrect suffixes suggests that the phonological constraints inherent in the Regular mapping are largely obeyed by the network.

Figure 21 shows the results of taking into account the correct mapping type for each lexical item, i.e., whether the MAPPING 2 relationship determined by the network is the same as that defined in the training set. Thus, for example, the Vowel Change graph shows the proportion of Vowel Change verbs in the training set that are indirectly mapped as a Vowel Change (again, irrespective of the correctness of the stem form). These graphs are

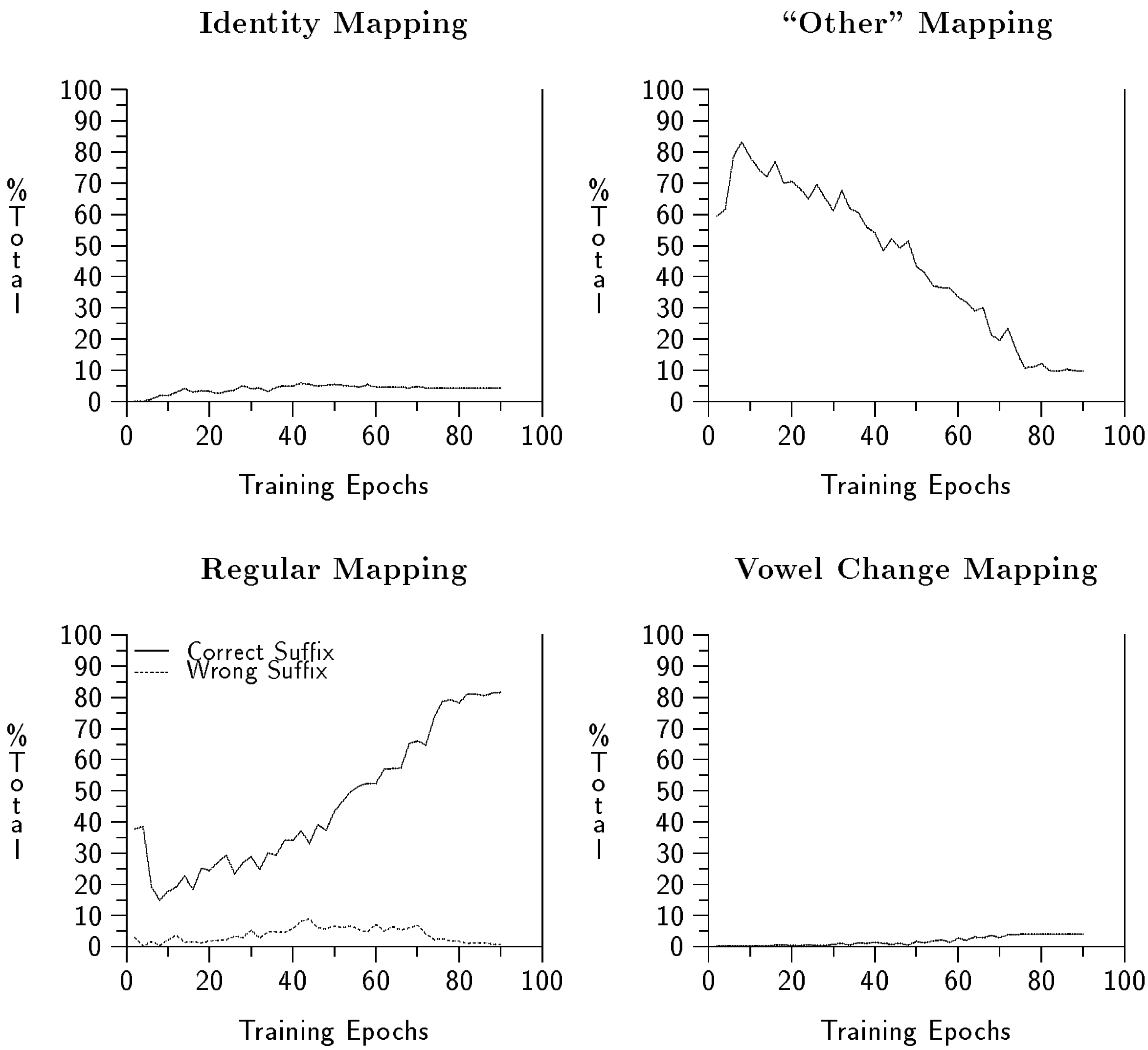


Figure 20: Percent of total trained stems whose past form is produced according to the different mapping types, measured based on the transformation of the (possibly incorrect) stem output of the network, and irrespective of the "correct" transformation for that lexical item. "Other" is all transformations that could not be classified into one of the three rule-based classes.

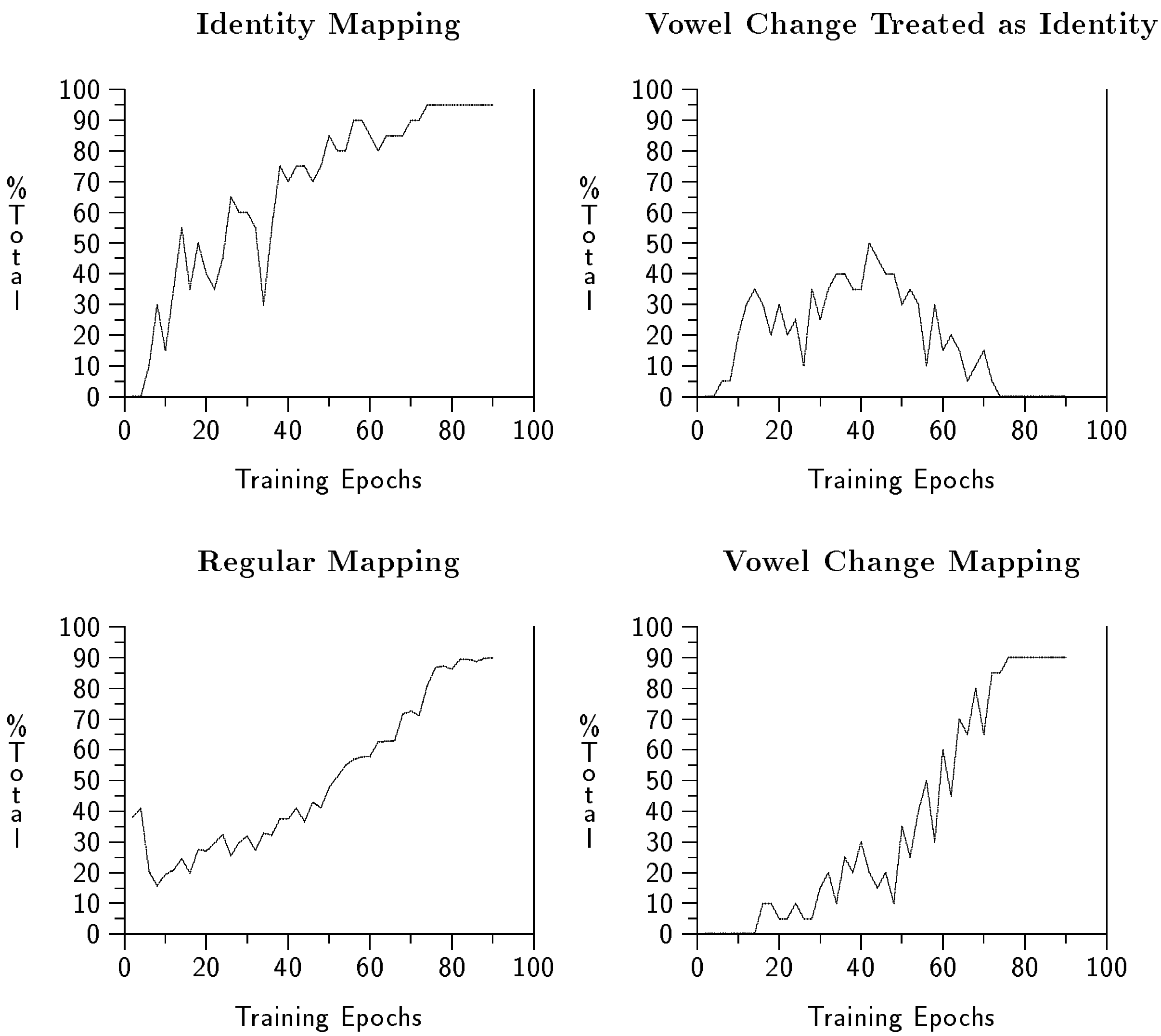


Figure 21: Proportion of mapping appropriate for each class. This is the same as in the previous figure, except that percentages are for each class as specified in the training set. The upper left panel shows the number of Identities actually treated as Identities. The upper right panel shows the proportion of Vowel Change verbs that were treated as Identities (no other interactions exist). The lower panels show the proportions of Regulars and Vowel Changes that were treated as such. Again, these proportions are calculated based on the transformation of the actual stem output, rather than the correct past.

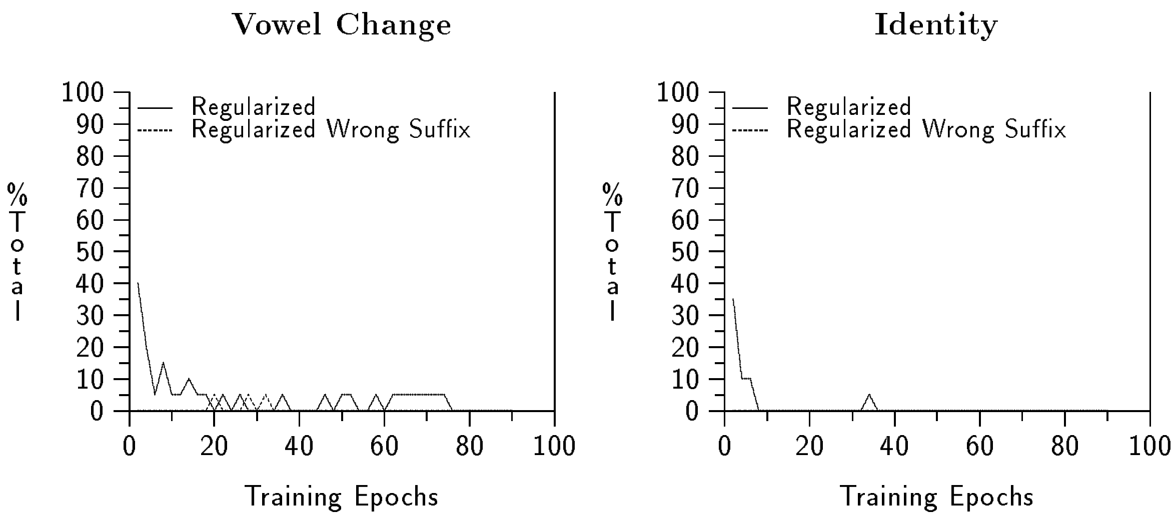


Figure 22: Overregularizations in the MAPPING-2 of irregulars, when transformation of the actual stem form is used for scoring.

thus analogous to the hit rate graphs of the previous section. The most striking feature of these graphs is that they show a very different acquisition profile than the hit rate graphs (see Figure 15). Here, the Regulars are shown to have a head start, although the Identities soon overtake them. The Vowel Change mapping is much worse than expected. The reason for this is shown in the fourth graph, where we see that a large proportion of the Vowel Changes are actually treated as Identity verbs. Thus the hit rates on the Vowel Change stems and pasts reported in the previous section are misleading because they display the hits on each form isolated from the behavior of the network on the alternate form. The current analysis reveals that the Vowel Change mapping is a very difficult one for this network. This would be expected by the low frequency of the transformation and the indirect way that the phonological conditioning must be induced.

Finally, this kind of analysis shows that early in learning, there is more overregularization going on than could be gleaned from the other analyses. Figure 22 shows the number of percent of each Vowel Change and Identities that are regularized.

10 Discussion

10.1 How does semantics influence the acquisition of adult forms?

The main variable we manipulated was similarity of meaning. The basic result from the simulations is that synonyms take longer to learn than words with highly differentiated meanings. This conforms to the results of Johnston & Slobin [1979], in that when meanings are similar, correct surface forms take longer to acquire. The explanation in terms of the model is that similar inputs tend to produce similar outputs. Thus it takes longer to differentiate these forms at the output level. This explanation thus also makes a prediction that the forms produced by the child for synonyms will be more similar to one another than randomly chosen elements of the child's vocabulary.

We also found a "squeeze" effect — when other words are being learned, surface forms of synonyms tend to become more similar. In their study, Johnston & Slobin evaluated children's competence with locative expressions at two time points, four months apart. They also found a decrement in performance at the later observation point for some children speaking a language with a highly differentiated vocabulary:

> The low Serbo-Croatian gain resulted in particular from the relatively large number of children who performed less well at Time II than at Time I. (p. 536)

We speculate that this decrement in performance could be due to a squeeze effect. During error correction learning, as the model differentiates the more disparate inputs, it does so at the cost of more similar inputs. We have characterized this as a competition for representational resources in the plan vector. In effect, synonyms become more like homonyms during the vocabulary burst. It is worth noting that in earlier work (Cottrell & Plunkett [1991]), we found a much greater squeeze effect with a much smaller vocabulary (fifty words). This may be more realistic in the case where semantics and surface forms are being learned simultaneously. Finally, in informal observation, we have noted that when these similar output forms were finally differentiated, they did so rapidly.

10.2 What is the source of children's production errors?

The model predicts that a fundamental source of children's production errors arise from the need to connect meaning to form. Difficulties arise in assigning distinctive phonological forms to very similar underlying meanings. Thus, the model predicts that children will have difficulty in acquiring words for highly similar concepts. We have characterized this as due to a competition for representational resources. The network is learning a plan for producing this sequence of phonemes. The network takes the easiest distinctions first, sacrificing the more difficult mappings until later. The model suggests that the most frequently practiced word of a group of words with very similar meanings will be the one used for the whole group.

This is thus an alternative explanation for the source of overextension errors to the standard account (Clark [1973]), where the errors are attributed to undifferentiated semantics. Clark hypothesizes that children will call all four-legged things "doggie" simply because their semantic structures do not yet differentiate four-legged objects such as cows and dogs. In contrast, our model suggests that the dog mapping "attracts" the cow mapping, since the inputs are similar. These explanations are not antithetical. As mentioned earlier, we do not take the assumption that the child has perfect versions of the input and target forms as an essential determinant of the behavior of the model. Rather, the model would respond on the basis of similarity of input representations even if the input forms were being learned at the same time as the mapping itself. Simply put, the model suggests that undifferentiated meanings may not be the only source of the error.

There will also be errors due to the architecture of the model. The obvious fact that the hardware for producing these words is shared between all words leads to the prediction that corrections to some forms will be detrimental to others. Thus we see W-shaped learning curves on many forms. This kind of effect was predicted by Stemberger [1992] in his discussion of connectionist learning models, and has been demonstrated most effectively by (Plunkett & Marchman [1991]).

In the next section, we discuss a third source of errors attributable to the structure of the output domain.

10.3 How does the set of forms themselves mold the child's developing vocabulary?

Plunkett [1993] has suggested that the idiosyncratic vocabularies of children arise because the children are having difficulties perceiving the boundaries between the words in the target language. The model presented here suggests an additional source of error. If one posits a single error-correcting system that is used for all words, then the words produced may reflect global regularities in the target domain, as our model demonstrated. Pseudowords that fit the syllabic structure of the adult vocabulary are acquired early, only later to be replaced by the correct forms as the system improves.

The kinds of forms produced reflect the frequencies of the forms in the target language. Our model showed a tendency to overshoot by overproducing the most frequent syllabic structure. The model then needed to recover from this by transforming the dominant syllabic structure into the less frequent ones. The restructuring involved producing forms that were not in the adult language, but were on trajectories between adult forms.

11 Conclusions

We have described a connectionist model of morphology acquisition in which input forms representing the meanings of words are mapped to sequences of outputs representing their phonological forms. The network is successful in producing appropriate forms, even though the input forms have a similarity structure that is independent of the similarity structure of the targets. Furthermore, the learning curves indicate a spurt-like acquisition profile. There is ample evidence for the spurt-like nature of vocabulary growth (McShane [1979]). It is unclear whether the acquisition of inflectional morphology in children shows a similar non-linear growth to that observed in the network.

We have also shown that the network is able to learn the relationship between different forms of the same verb, in spite of the fact that it is only exposed to these relations indirectly. In particular, we found that stem/past-tense pairs honored an appropriate form relationship, even when the stem's phonological form did not conform to the target signal. In effect, the model has acquired a generative capacity for inflecting verbs given that it knows one member of the paradigm. The model predicts that children are better at generalising from past tense forms to stems than *vice versa*. Further analysis is needed to investigate what modifications must be made to the model in order to achieve better generalization in the structured input case. The results of Bartell, Cottrell & Elman [1991] suggest that a layer of hidden units between the recurrent hidden layer and the output will aid generalization based on output similarity.

The analysis of the influence of input and target similarity on the acquisition of phonological form suggests some radical predictions. Children's non-adult forms may be a result of blending words from the same semantic category. Looked at another way, words are distorted by their neighbors in a semantic class. The effects of similarity at the phonological level suggest that children will initially over- and under-shoot the correct proportion of syllablic structures for their language. During the correction phase, they will produce forms that do not belong to the syllabic structure of the language if these forms are between the most common form in the language and other forms. Finally, the model suggests that during the vocabulary burst, synonyms will be forced to be near-homonyms.

References

Barrett, M. D. [1978], "Lexical development and overextention in child language," *Journal of Child Language* 5, 205–219.

Bartell, B., Cottrell, G. W. & Elman, J. [1991], "The role of input and target similarity in assimilation.," *Proceedings of the Thirteenth Annual Cognitive Science Society Conference, Chicago, IL*, Hillsdale, NJ.

Bybee, J. & Slobin, D. I. [1982], "Rules and schemas in the development and use of the English past tense," *Language* 58, 265–289.

Chauvin, Y. [1988], "Symbol Acquisition in Humans and Neural (PDP) Networks," University of California, San Diego, PhD Thesis.

Clark, E. V. [1973], "What's in a word? On the child's acquisition of semantics in his first language," in *Cognitive development and the acquisition of language*, T. E. Moore, ed., Academic Press, New York.

Corina, David P. [1991], "Towards an understanding of the syllable : Evidence from linguistic, psychological, and connectionist investigations of syllable structure," *Ph.D. Thesis*, La Jolla, CA.

Cottrell, G. W. & Plunkett, K. [1991], "Learning the past tense in a recurrent network: Acquiring the mapping from meanings to sounds," *Proceedings of the Thirteenth Annual Cognitive Science Society Conference, Chicago, IL*, Hillsdale, NJ.

Dale, P. S. [1972], *Language development: Structure and Function.*, Dryden Press, New York.

Duda, R. O. & Hart, P. E. [1973], *Pattern Classification and Scene Analysis*, Wiley, New York.

Elman, J. L. [1990], "Finding structure in time," *Cognitive Science* 14, 179–211.

Gasser, M. & Lee, C. [1990], "Networks that learn about Phonological Feature Persistence," *Connection Science* 2, 265–278.

Gasser, Michael [1994], "Acquiring receptive morphology: A connectionist model," *Proceedings of the 32nd Annual Meeting of the Associaion for Computational Linguistics*, San Fancisco, CA.

Hare, M. [1990], "The Role of Similarity in Hungarian Vowel Harmony: a Connectionist Account," *Connection Science* 2, 123–150 .

Hare, M., Corina, D. & Cottrell, G. W. [1989], "A connectionist perspective on prosodic structure," *Proceedings of the 15th Annual meeting of the Berkeley Linguistics Society*, Berkeley, CA.

Jacobs, R., Jordan, M. I., Nowlan, S. J. & Hinton, G. E. [1991], "Adaptive mixtures of local experts.," *Neural Computation* 3, 79–87.

Jespersen, O. [1925], *Language*, Holt, Rinehart and Winston, New York.

Johnston, J. R. & Slobin, D. I. [1979], "The development of locative expressions in English, Italian, Serbo-Croation and Turkish," *Journal of Child Language* 6, 529–545.

Jordan, M. [1986], "Serial order: A parallel distributed processing approach," Institute for Cognitive Science. University of California, Technical Report #8604, San Diego.

Jordan, M. I. & Rumelhart, D. E. [1992], "Forward models: Supervised learning with a distal teacher.," *Cognitive Science* 16, 307–354.

Jusczk, Peter W. [1992], "Developing phonological categories from the speech signal.," in *Phonogical development*, C. Ferguson, L. Menn & C. Stoel-Gammon, eds., York Press, Timonium, MD, 17–64..

Knapp, A. G. & Anderson, J. A. [1984], "Theory of categorization based on distributed memory storage," *Journal of Experimental Psychology: Learning, Memory and Cognition* 10, 616–637.

MacWhinney, B. & Leinbach, J. [1991], "Implementations are not conceptualizations: Revising the verb learning model," *Cognition* 40, 121–157.

Marchman, V. [1988], "Rules and regularities in the acquisition of the English past tense," *Center for Research in Language Newsletter* 2.

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.

McShane, J. [1979], "The development of naming," *Linguistics* 17, 879–905.

Movellan, J. R. [1990], "Contrastive Hebbian learning in interative networks," Unpublished manuscript, Department of Psychology, Carnegie Mellon University.

Pinker, S. & Prince, A. [1988], "On language and connectionism: Analysis of a parallel distributed processing model of language acquistion," *Cognition* 28, 73–193.

Plaut, D. C. & Shallice, T. [1993], "Deep Dyslexia: A Case Study of Connectionist Neuropsychology.," *Cognitive Neuropsychology* 10, 377–500.

Plunkett, K. [1993], "Lexical segmentation and vocabulary growth in early language acquisition," *Journal of Child Language* 20, 43–60.

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

Plunkett, K. & Marchman, V. [1993], "From Rote Learning to System Building: Acquiring Verb Morphology in Children and Connectionist Nets," *Cognition* 48, 21–69.

Plunkett, K. & Sinha, C. G. [1992], "Connectionism and Developmental Theory," *British Journal of Developmental Psychology* 10, 209–254.

Rumelhart, D. E. & McClelland, J. L. [1986], "On learning the past tense of English verbs," in *Parallel distributed processing: Explorations in the Microstructure of Cognition, #2: Psychological and Biological Models*, J. L. McClelland, D. E. Rumelhart & PDP Research Group, eds., MIT Press, Cambridge, MA, 216–271.

Stemberger, J. [1990], *The acquisition of morphology: Analysis of symbolic model of language acquisition*, University of Minnesota, Unpublished manuscript.

Stemberger, J. [1992], "A connectionist view of child phonology: phonological processing without phonological processes.," in *Phonogical development*, C. Ferguson, L. Menn & C. Stoel-Gammon, eds., York Press, Timonium, MD.

Vihman, M., Macken, M., Miller, R., Simmons, H. & Miller, J. [1985], "Form babbling to speech: a re-assessment of the continuity issue," *Language* 61, 397–443.