Relating Comprehension and Production in the Acquisition of Morphology*

Michael Gasser
Indiana University

Abstract

Most theories of language processing and acquisition make the assumption that perception and comprehension are related to production, but few have anything say about how. This paper describes a performance-oriented connectionist model of the acquisition of morphology in which production builds on representations which develop during the learning of word recognition. Using artificial language stimuli embodying simple suffixation, prefixation, and template rules, I demonstrate that the model generalizes to novel combinations of roots and inflections for both word recognition and production. I argue that the capacity of connectionist networks to develop intermediate distributed representations which not only enable the solving of the task at hand but also facilitate another task offers a plausible account of how comprehension and production come to share phonological knowledge as words are learned.

Introduction

Language learners must acquire both the ability to comprehend language and the ability to produce language. While the extent of the relationship between these abilities is still controversial (see Harris et al., 1995 and Elbers, 1995 for two recent examinations), it almost goes without saying that there is a relationship. Perhaps since linguistics has a strongly production-oriented flavor to it, modeling, even connectionist modeling, has tended to focus on production. The child, on the other hand, starts with perception and comprehension; forms cannot normally be produced until they have been heard.

This paper considers the question of how perception and comprehension on the one hand and production on the other are to be related in a performance-oriented theory of language acquisition and, in particular, the question of how production could be learned at all. The focus is on the acquisition of morphology, but this can only be addressed in the context of the acquisition of phonology. I describe a partially implemented computational model of the acquisition process in which distributed phonological representations provide knowledge which is shared by receptive and productive processing.

*Paper presented at the Groningen Assembly on Language Acquisition, Groningen, September 1995

Word Comprehension and Production in Language Acquisition

Let us consider the acquisition of word comprehension and production from the perspective of what information is available to the child. Word comprehension means mapping auditory[1] form onto meanings. This task can be viewed as an example of *supervised* learning: under at least some circumstances, the child has access to a “teacher” to provide the correct response. That is, because language addressed to children tends to refer to the here-and-now (Snow, 1977), the referent of the word is available in the context and often pointed to in one way or another. The evidence the child receives is far from perfect — words are not always isolable in the input stream, and the precise set of semantic features singled out by the word is never actually indicated (Quine, 1960) — but in a sense the child is “taught” to understand words. For word recognition there is a target.

The same cannot be said for word production. Children who produce an utterance with some communicative intent have no target to guide them. Consider the first opportunity to produce a new word, that is, the intent to refer to a type of object or relation for the first time. For this task, there is no target whatsoever, no set of articulatory gestures, albeit incomplete, provided by the environment in response to the child’s intent. The child may of course produce something in response to this intent, If this is interpreted as wrong, there might be correction: “no, that’s not a potato; that’s oatmeal.” However, this sort of feedback in production is less useful than what is available in comprehension. First, the adult hearer has no way of knowing what the original intent of the speaker was: perhaps the intended meaning concerns the similarity of oatmeal and potatoes; perhaps the child knows full well what potatoes are and has simply not noticed that this is something else. Second, and more importantly for our purposes, the output of production is a sequence of articulatory gestures; correction would seem to require reference to these for the child to make direct use of it. But correction is never of the form “no, you should have first rounded your lips like this”, and if it were, it would of course be unintelligible to the child. For production, the child faces the *credit assignment problem*: “something was apparently wrong with what I did, but what?” Unlike comprehension, production cannot be viewed as a supervised task.

How, then, can production be learned at all? The only real possibility is through *reinforcement*, rather than strictly supervised, learning (Sutton, 1992). When the learner is right and is told so, reinforcement learning is identical to supervised learning: the learner should do the same thing under the same circumstances the next time. When the learner is wrong and is told so, however, reinforcement learning provides no information about what precisely was wrong. However, it is hard to imagine that the child learns everything there is to know about word production in this trial-and-error sort of fashion. This would amount to guessing the sequence of articulatory gestures for all novel meanings and then attempting to revise each guess on the basis of the information that the result was somehow wrong. The problem is greatly simplified if we assume that there is transfer from comprehension to production. That is, as the child learns associations between form and meaning, the child also learns the reverse associations between

[1] For convenience, I will refer to the input to perception as “auditory” or “acoustic,” but this is not meant to exclude the possibility of visual input, as one would have for a signed language.

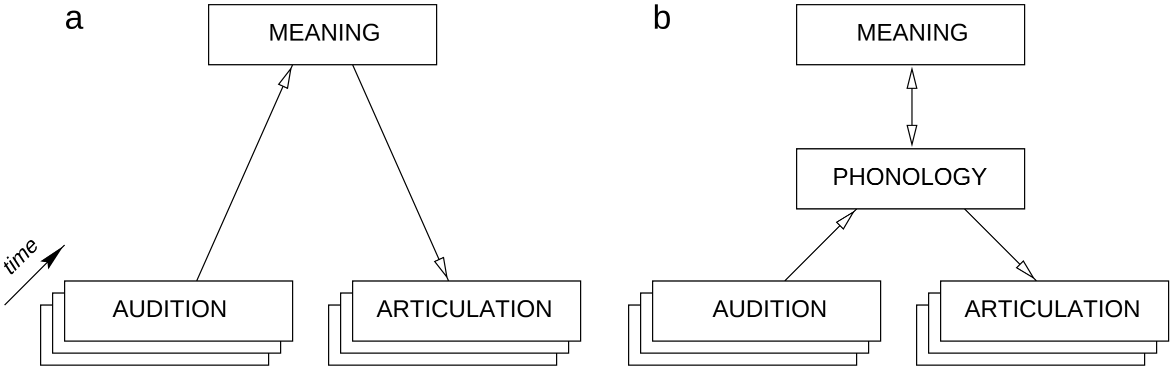

Figure 1: Perception and Production of Words: Two Possibilities

meaning and form. This alternative fails, however, if there is nothing linking auditory and articulatory form; learning to map an auditory form onto a meaning says nothing about how that meaning should be mapped onto an articulatory pattern. Auditory and articulatory form must share something intermediate; two-way associations could then be learned between meaning and these shared intermediate representations.

These relationships are illustrated in Figure 1. In Figure 1a, we see the situation when there is no sharing; production has no way to benefit from perception and comprehension. In Figure 1b, there is an intermediate level at which receptive and productive performance make use of the same representations. I will refer to this level as that of **intermediate phonological representations (IPRs)** both "phonological" and "representation" have non-standard uses here. From the perspective of perception and comprehension, IPRs are a level at which the mapping from acoustic/auditory form to meaning is mediated. From the perspective of production, the IPR level divides the learning task into two components. One, the mapping between meaning and IPRs, is easy to solve if we can solve the reverse mapping for perception and comprehension. The other, the mapping from IPRs to articulation, does not benefit, at least not directly, from perceptual learning. However, in order for the learner to make use of the meaning-to-IPR associations, this other mapping must be partly in place. Note that given IPRs, correction of the form "no, it's not a potato, it's oatmeal" becomes useful to the child. The auditory input corresponding to "oatmeal" gets translated into an IPR for this word, which can then be associated with semantic features tied to the referent. The IPR-to-articulation component is bypassed altogether.

Of course there is nothing novel in arguing that phonology is shared by comprehension and production. What is new here is the argument for why this must be so and, in what follows, for the nature of the phonological representations, which bear few resemblances to those of traditional phonology.

For the sake of argument, let us assume that for a particular language IPRs correspond to syllables. The IPR-to-articulation learning task, then, consists in learning how abstract syllable representations are to be translated into sequences of articulatory gestures. This is just the learning of articulatory phonology. A tentative solution to this mapping could be found by the learner prior to the learning of semantics and morphology, that is, during babbling. Of course, babbling is not a supervised task; the child does not have the advantage of explicit targets. From the perspective of the present account, babbling may be seen as follows. Given a particular input, possibly the result of some perceived linguistic input, the babbler attempts to produce a sequence of articulatory gestures which reproduces it. She has access to the acoustic/auditory consequences of what she does, but the only feedback she gets is whether her articulatory sequence is right or wrong (and perhaps by how much). Thus this is an instance of reinforcement learning. The main point here is that the acquisition of word production becomes tractable if divided into two tasks, one, the mapping of meaning onto phonology, an instance of supervised learning, and the other, the mapping of phonology onto articulation through reinforcement learning.

On this account, then, the learning of phonology takes place in three overlapping "stages." These are illustrated in Figure 2. During the first several months of life, the child focuses on the initial learning of IPRs on the basis of auditory input (Figure 2a). This is an example of *unsupervised learning*; there is no teacher at all. The learner is simply attempting to find regularity in the input patterns. When the child begins to produce language-like sounds in the middle of the first year, the learning of the IPR-to-articulation associations can begin (Figure 2b). For a given utterance, the input is a legitimate IPR. Based on the learner's current state of knowledge, this results in an articulatory sequence. Using feedback from her own perceptual system (via the dashed arrow in the figure), the child judges the articulation to be right or wrong to some degree. Since this feedback cannot specify precisely how it is wrong, this is reinforcement learning. The result of this arduous process is a system which can sound roughly as it wants to sound.

But phonology is also constrained by the lexicon and by morphology; the contrasts that matter in the target language only become evident when morphemes are learned. During a subsequent phase, which may begin long before babbling ends, the first words are learned (Figure 2c). This involves both IPR-to-meaning learning and also the modification of the associations among IPRs and also possibly the auditory-to-IPR associations. Because this results in modifications to the IPRs themselves, the articulatory learning that went on during babbling must also continue during this phase. Thus the IPR-to-articulation associations continue to be modified.

Desiderata

What, then, should we expect of a model of morphology acquisition?

1. We expect phonological representations which are shared by receptive and productive processes, representations which arise as the system is exposed first to utterances and later to words paired with their meanings.

2. We expect these shared phonological representations to enable generalization from comprehension to production. Given a new auditory input, the learning system should yield a representation suitable for production as well as comprehension.

   (a) When a novel word is presented, the phonological representation should be interpretable by the production system as a sequence of articulatory gestures

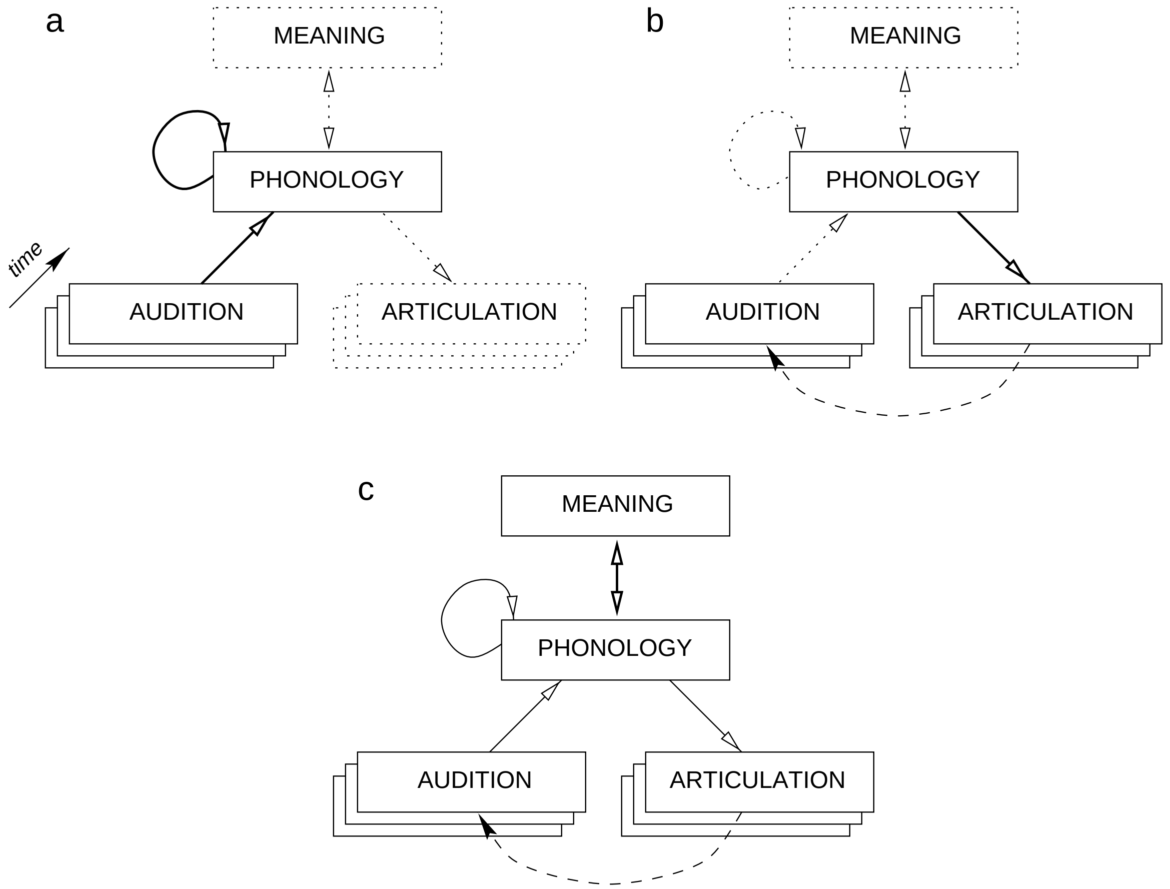

Figure 2: Stages in the Learning of Phonology

whose acoustic consequences approximate the original acoustic input.

(b) When a morphological "rule" is learned during comprehension, this should apply to production as well. That is, once comprehension training has taken place, the presentation of a novel combination of morpheme meanings, when input to production, will yield an appropriate articulatory output.[2]

In the next section, I describe a partially implemented computational model which embodies these features.

A Model: MCNAM

The Modular Connectionist Network for the Acquisition of Morphology (MCNAM) is a relatively simple architecture which has been shown in previous work (Gasser, 1994a) to be capable of learning morphological rules of the following types: suffixation, prefixation,

[2] There is a further possibility for the learning of morphology, namely, that analysis into constituent morphemes takes place as the child listens to her own production of a memorized word (Elbers, 1995). While this complicates the picture presented here somewhat, it is still in agreement with the basic account of how receptive and productive learning are related.

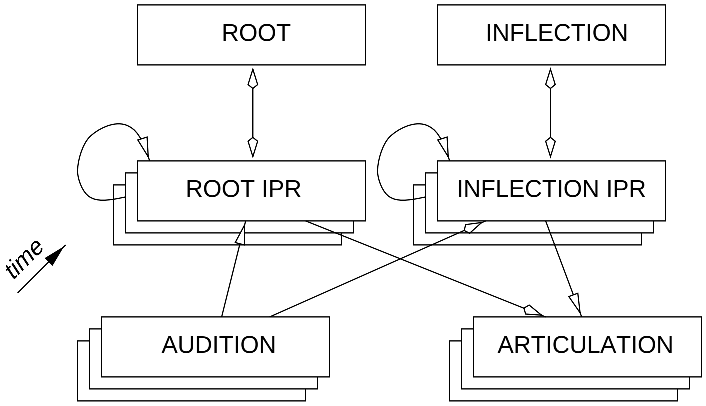

Figure 3: Architecture of MCNAM

infixation, mutation, template, and deletion.[3]

The network consists of separate, though connected, modules for receptive (perception and comprehension) and productive performance. Further the receptive component consists of separate modules responsible for recognition of the root and for the grammatical morphemes in a word (Gasser, 1994b). Each module in the network is a form of simple recurrent network (Elman, 1990), a feedforward network augmented with some recurrent (feedback) connections.

The overall architecture of the model is shown in Figure 3. Boxes represent clusters of connectionist units, and arrows represent complete connectivity between clusters. There are recurrent connections on the hidden-layer (IPR) units. That is, in receptive mode, each hidden-layer unit receives input not only from all of the AUDITION units but also from the hidden layer units in the same cluster, including the unit itself. The values that are passed from one hidden layer unit to another are those from the previous time step; there is a delay on the recurrent connections.

The production module is actually somewhat more complicated than what is shown in the figure. It consists of two networks which are run and trained separately, one connecting the MORPHEME (ROOT and INFLECTION) clusters to the IPR units, the other connecting the IPR cluster to the ARTICULATION units. Each of these subnetworks is actually somewhat more complicated that what is shown in the figure; it has recurrent connections and an additional hidden layer.

The network is trained as follows. An initial phase corresponds to the pre-lexical learning of phonology. It only very approximately models the two stages described above. The unsupervised learning of initial IPRs based on auditory information is avoided altogether. Instead the connections joining the AUDITION and IPR clusters are assigned initial random weights, and a set of input syllables is presented to the untrained network. Following each syllable, the pattern on the IPR cluster is saved as a representation of that syllable.

Next the IPR-to-ARTICULATION subnetwork is trained on a subset of the syllables

[3] Reduplication and metathesis rules are not learnable by the model in the form described in this paper. They appear to require a hierarchical architecture in which prosodic levels (syllable, foot, etc.) are represented explicitly in the network.

used to generated the IPR patterns in the AUDITION-to-IPR network. The remaining one-third of the forms are tested to determine whether the IPR-to-ARTICULATION mapping that has been learned generalizes to novel input syllable sequences. While learning at this stage should in fact be on the basis of reinforcement, for the simulations I make use of the more powerful supervised learning. The point is simply to show that the IPRs contain the sort of information that supports articulation.

Next the AUDITION-to-MORPHEME subnetwork is trained. The inputs consist of sequences of "auditory" segments representing inflected words, and the outputs are ROOT and INFLECTION patterns, that is, a ROOT vector with a single unit on, the remaining units off, and an INFLECTION vector with a single unit on, the remaining units off. This network is trained on two-thirds of the words and tested for generalization on the other one-third. Training is supervised.

Finally, both the MORPHEME-to-IPR and IPR-to-ARTICULATION subnetworks are trained, with combinations of roots and inflections as MORPHEME inputs and IPR (syllable) sequences as outputs for the first task and as inputs for the second, and ARTICULATION sequences as outputs for the second task. Again two-thirds of the possible words are used for training with the rest set aside for testing the network. If the IPR representations learned for comprehension are suitable for production, the production network should produce appropriate IPR sequences for morpheme combinations which it has not been trained on, and these IPR sequences should be interpretable as appropriate ARTICULATION sequences. The main point of the simulations described below is to demonstrate this generalization capacity.

Simulations

In each of the three sets of experiments described here, the network is trained on two-thirds of the possible patterns and tested on the remaining third. Of interest in each case is whether the network generalizes, that is, whether it responds appropriately to novel test patterns. Generalization indicates that the network has actually learned the rule behind the forms.

Stimuli The network was trained on artificial language stimuli. Words in the language were composed of sequences of two or more CV syllables. The consonant inventory was [*p*, *b*, *f*, *m*, *t*, *d*, *s*, *n*, *k*, *g*, *x*, *ŋ*], the vowel inventory [*a*, *i*, *u*]. There were three separate morphological "rules": suffixation, prefixation, and template; each network was trained on only one of these. For the affixation rules, 12 roots each of one and two syllables were generated randomly. The affixes themselves consisted of *fi*, *ni*, and *ku*, prefixed in one case, suffixed in the other. Thus for the prefixation case, the three forms for the root *bapu* were *fibapu*, *nibapu*, and *kubapu*. For the template rules, 24 two-consonant roots were generated randomly. The inflected forms then consisted of two successive syllables with the same vowel. Thus the three forms for the root *bx* were *baxa*, *bixi*, and *buxu*.

"Auditory" inputs consisted of sequences of segments, each containing values for seven gross acoustic features. "Articulatory" output also consisted of sequences of segments, but these were based loosely on the gestures of Articulatory Phonology (Browman

& Goldstein, 1986). Thus, while these inputs and outputs were far from authentic, they were quite unlike each other in character.

With 24 roots and three inflections, there were always 72 possible words. In each case the network was trained on 48 of these, two of the three forms for each root, and tested on the remaining 24.

Generalization performance was evaluated by determining which of the possible responses was closest to the network's output pattern in Euclidean distance.

Production from IPRs The first experiment was concerned with initial learning on the IPR-to-ARTICULATION connections.

First, an AUDITION-to-IPR network was set up with random weights. Next this network was presented all 36 possible syllables in the language. At the end of each syllable the pattern on the IPR cluster was saved. Note that this network was *not* trained; the IPR patterns were those resulting from the initial weights on the AUDITION-to-IPR connections and the recurrent IPR-to-IPR connections.

Next the IPR-to-ARTICULATION network was trained on two-thirds of the syllables. The inputs in each case were the IPR patterns themselves. Training was supervised (rather than via reinforcement, as would be the case with the child). Thus for the syllable *gu*, the input was the pattern that had appeared on the IPR cluster following presentation of the sequence of "auditory" patterns corresponding to that syllable to the AUDITION cluster. However, each syllable pattern was presented for four time steps as the articulatory output and target changed. This was necessary because each syllable corresponded to a sequence of four "gesture" patterns on the ARTICULATORY cluster.

Following training, the IPR-to-ARTICULATION subnetwork was tested on the remaining one-third of the syllables. After considerable training, the network correctly generated 58% of the "articulatory" segments in the test syllables; that is, for 58% of the segments the network's output was closer to the correct segment than to any other. Since there were 44 possible segments in all, this performance is far above chance. While only three of the 12 test syllables were produced perfectly, most of the errors on other syllables were reasonable ones (*gu* → *ŋu*, *du* → *gu*, *ni* → *mi*, *sa* → *ga*, *ta* → *ka*).

These results show that the distributed syllable representations from the AUDITION-to-IPR network, even when it has not been trained at all, contain the sort of information which permits generalization on the production task, the task of transforming these syllable representations to articulatory sequences.

Learning to Recognize Words The next two phases of training involved the learning of polymorphemic words. First the recognition subnetwork was trained in a supervised fashion to recognize the words. Separate networks were trained on the three rules, prefixation, suffixation, and templates. Inputs consisted of "auditory" sequences; targets consisted of the appropriate ROOT and INFLECTION.

Generalization results were as shown in Table 1. Note that chance performance would be 4% for roots and 33% for inflections.

Performance on root recognition for the template rule is quite low (though still far

TYPE	ROOT	INFLECTION
suffixation	75%	92%
prefixation	71%	100%
template	25%	100%

Table 1: Performance on Test Words for Recognition

TYPE	MORPHEME-to-IPR	IPR-to-ARTICULATION
suffixation	92%	74%
prefixation	87%	70%
template	65%	47%

Table 2: Performance on Test Words for Production

above chance), but it rises to 67% when we include output roots which differ by only one phonetic feature, for example, *kp* for *kb*.

The results indicate clearly that the network is capable of learning to recognize the root and inflection resulting from simple affixation and template rules.

Learning to Produce Words Next the hidden layer following each syllable of the input words was saved for training the production subnetwork. As described above, the two portions of the production network were trained separately. As with recognition, separate networks were trained for the three morphological rules, and training is supervised. The MORPHEME-to-IPR network took constant patterns representing ROOT and INFLECTION as inputs and sequences of IPR patterns, one for each syllable, as targets. The IPR-to-ARTICULATION network took sequences of IPR patterns as inputs and sequences of "articulatory" segments as targets.

Generalization performance for both portions of the production network is shown in Table 2. Chance performance was 2.8% for the MORPHEME-to-IPR task and 2.3% for the IPR-to-ARTICULATION task.

Again performance is always far better than chance. The results indicate that the phonological representations learned during word recognition embody the structure that supports the learning of production.

Interestingly, training the production component with the IPRs from the word recognition appears to yield better performance than results when a network is trained on the MORPHEME-to-ARTICULATION task directly. A recurrent network with the same number of hidden units as contained in the IPRs (50), when trained on this task with the suffixation rule, achieved a performance level of only 58%. By comparison, if we combine the performance on the two production components using the intermediate IPRs during training, generalization is approximately 70%.

Conclusions

In this paper, I have (1) argued why and how receptive and productive performance might relate to each other in the acquisition of the lexicon and morphology and (2) shown how connectionism provides a way in which this relationship might be implemented.

Production must build on perception and comprehension. What is required are *shared representations*, derived from perception and comprehension but usable also by production. These representations should support the learning of production by providing a way for the process to be broken down into two more manageable tasks, learning to translate semantic input into phonological representations and learning to translate phonological representations into articulatory gestures.

One of the appeals of connectionist models is their ability to solve specific tasks by developing internal representations which at the same time embody the internal structure that is necessary to solve other tasks. I have demonstrated that a particular type of modular simple recurrent connectionist network, in learning to map pseudo-auditory input sequences onto the morphemes, recodes the inputs as sequences of distributed patterns, which in turn support the learning of word production.

A further benefit of this model, one not explored at all in this paper, is that it makes very specific predictions about the sorts of comprehension and production errors that appear in the learning of words and of morphology.

Despite numerous gaps in the model itself, in particular, the absence of reinforcement and unsupervised learning, the present account shows promise. Stated simply, production is learnable because comprehension is learnable and because there is phonology to tie the two together.

References

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.

Elbers, L. (1995). Production as a source of input for analysis: evidence from the developmental course of a word-blend. *Journal of Child Language*, *22*, 47–71.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Gasser, M. (1994a). Acquiring receptive morphology: a connectionist model. *Annual Meeting of the Association for Computational Linguistics*, *32*, 279–286.

Gasser, M. (1994b). Modularity in a connectionist model of morphology acquisition. *Proceedings of the International Conference on Computational Linguistics*, *15*, 214–220.

Harris, M., Yeeles, C., Chasin, J., & Oakley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language*, *22*, 1–18.

Quine, W. V. O. (1960). *Word and Object.* MIT Press, Cambridge, MA.

Snow, C. E. (1977). Mothers' speech research: from input to interaction. In Snow, C. E. & Ferguson, C. (Eds.), *Talking to Children: Language Input and Acquisition*, pp. 31–49. Cambridge University Press, Cambridgem MA.

Sutton, R. S. (Ed.). (1992). *Reinforcement Learning.* Kluwer Academic, Boston.