STUDIES IN LANGUAGE COMPANION SERIES (SLCS)

The SLCS series has been established as a companion series to STUDIES IN LANGUAGE, International Journal, sponsored by the Foundation "Foundations of language".

Volume 41

Michael Darnell, Edith Moravcsik, Frederick Newmeyer,
Michael Noonan and Kathleen Wheatley (eds)

Functionalism and Formalism in Linguistics
Volume I: General Papers

FUNCTIONALISM AND FORMALISM IN LINGUISTICS

VOLUME I: GENERAL PAPERS

Edited by

MICHAEL DARNELL
EDITH MORAVCSIK
FREDERICK NEWMEYER
MICHAEL NOONAN
KATHLEEN WHEATLEY

1999

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

Emergent Language

Brian MacWhinney
Carnegie Mellon University

Abstract

Recent work in language acquisition has shown how linguistic form emerges from the operation of self-organizing systems. The emergentist framework emphasizes ways in which the formal structures of language emerge from the interaction of social patterns, patterns implicit in the input, and pressures arising from general aspects of the cognitive system. Emergentist models have been developed to study the acquisition of auditory and articulatory patterns during infancy and the ways in which the learning of the first words emerges from the linkage of auditory, articulatory, and conceptual systems. Neural network models have also been used to study the learning of inflectional markings and basic syntactic patterns. Using both neural network modelling and concepts from the study of dynamic systems, it is possible to analyze language learning as the integration of emergent dynamic systems.

If you spend some time watching the checkout lines at a supermarket, you quickly find that the number of people queued up in each line is roughly the same. At peak times, you may find five or six people in a line waiting to check out. At slower times, lines have only two or three waiting. There is no fixed rule governing this pattern. Instead, the rule that equalizes the number of shoppers in the various lines emerges from other basic facts about the goals and behavior of shoppers and supermarket managers. This simple idea of emergence through constraint satisfaction is currently being invoked as a central explanatory mechanism in many areas of cognitive science and neuroscience.

Given the often effortless nature of language use, the idea of viewing verbal behavior as an emergent process seems particularly attractive. We can observe speakers carrying on conversations on cellular phones while driving their cars in rush hour traffic, and we can find accomplished seamstresses creating elaborate embroidery while conversing fluently. It is not only adult language processing that seems effortless; language learning in children also appears natural and painless.

Despite these appearances, when linguists look at language learning and processing, they find complex rules, categories, and symbols. How can we reconcile these divergent perceptions? One possible reconciliation calls into question the extent to which language learning and processing actually function in obedience to an explicit set of formal rules. According to this new view of language learning and processing, the behaviors that we tend to characterize in terms of rules and symbols are in fact emergent patterns that arise from the interactions of other less complex or more stable underlying systems. I will refer to this new viewpoint on language learning and processing as "emergentism".

Proponents of functional linguistics have often spoken of grammar as an emergent property of features of discourse (Du Bois 1987; Hopper & Thompson 1984), contrasting their functional analysis with formalist approaches to grammar. The idea that grammar can emerge from discourse is fundamental to the debate between functionalism and formalism in linguistics and psycholinguistics. However, the emergence of grammar from discourse is only one aspect of a much broader emergentist vision of the shape of human language. The shape of human language is also tightly governed by the physiology of the vocal apparatus, the nature of the auditory system, and the development and decay of the many cognitive systems that manage the processing of language. When we consider these various additional constraints on the emergent shape of language, we reach a broader characterization than that offered in functionalist accounts that look only at discourse pressures.

Emergentist accounts have been formulated for a wide variety of linguistic phenomena, ranging from segmental inventories, stress patterns, phonotactic constraints, morphophonological alternations, lexical structures, pidginization, second language learning, historical change, on-line phrase attachment, and rhetorical structures. Formalisms that have been used to analyze the emergent nature of these forms include connectionist networks, dynamic systems theory, neuronal differentiation models, classifier systems, production-system architectures, Bayesian models, Optimality Theory, and corpora studies.

The basic notion underlying emergentism is simple enough. Consider the hexagonal shape of the cells in a honeycomb. There is nothing in the genetic makeup of the honey bee that determines that each cell in the honey comb should take on the form of a hexagon. However, when circles are packed together, it turns out that packing distance is minimized when each circle has six neighbors. This same principle also applies in three dimensions to spheres. When

the fluid in these six neighboring honey cells is tightly compressed against its neighbors, a hexagonal shape emerges. No rules are needed to control the shape of each individual cell of the honeycomb; instead this form emerges from the interaction of hundreds of small units. Nature is replete with examples of this type of formal emergence. The form of beaches and mountain ridges, the geometry of snowflakes and crystals, the appearance of *fata morgana*, and the movement of the jet stream in the air and the Gulf Stream in the sea — all of these patterns arise from interactions of physical principles with constraints imposed by physical bodies. Even in the biological world, much of our somatic form is emergent, whether it be the patterns of stripes on the tiger, the formation of teeth into a uniform bite, the structuring of enzymes to catalyze organic reactions, or our patterns of fingerprints and hair formations.

1. Basic Assumptions

In this paper, we will explore three levels of emergent linguistic structure. The first level involves the acquisition of basic lexical structures in small areas of cortex called "local maps". The second level involves the interaction between lexical structures in terms of "lexical groups". The third level involves the processing of syntactic information across longer neural distances in "functional neural circuits". We will examine how linguistic form emerges from the interaction of these three levels of neurolinguistic processing.

2. Principles of Neural Networks

Connectionist models are implemented in terms of artificial neural networks. Neural networks that are able to learn from input are known as "adaptive neural networks". The architecture of an adaptive neural network can be specified in terms of eight design features:

1. *Units*. The basic components of the network are a number of simple elements called variously "neurons", "units", "cells", or "nodes". In Figure 1, the units are labeled with letters such as "x_1".
2. *Connections*. Neurons or pools of neurons are connected by a set of pathways which are typically called "connections". In most models, these connections are unidirectional, going from a "sending" unit to a "receiving"

unit. This unidirectionality reflects the fact that neural connections also operate in only one direction. The only information conveyed across connections is activation information. No signals or codes are passed. In Figure 1, the connection between units x_1 and y_1 is marked with a thick line.

3. *Patterns of connectivity*. Neurons are typically grouped into pools or layers. Connections can operate within or between layers. In some models, there are no within-layer connections; in others all units in a given layer are interconnected. Units or layers can be further divided into three classes:
   a. *Input units* which represent signals from earlier networks. These are marked as "x" units in Figure 1.
   b. *Output units* which represent the choices or decisions made by the network. These are marked as "z" units in Figure 1.
   c. *Hidden units* which represent additional units juxtaposed between input and output for the purposes of computing more complex, nonlinear relations. These are marked as "y" units in Figure 1.
4. *Weights*. Each connection has a numerical weight that is designed to represent the degree to which it can convey activation from the sending unit to the receiving unit. Learning is achieved by changing the weights on connections. For example, the weight on the connection between x_1 and y_1 is given as .54 in Figure 1.
5. *Net inputs*. The total amount of input from a sending neuron to a receiving neuron is determined by multiplying the weights on each connection to the receiving unit times the activation of the sending neuron. This "net input" to the receiving unit is the sum of all such inputs from sending neurons. In Figure 1, the net input to y_1 is .76, if we assume that the activation of x_1 and x_2 are both at "1" and the x_1y_1 weight is .54 and the x_2y_1 weight is .22.
6. *Activation functions*. Each unit has a level of activation. These activation levels can vary continuously between "0" and "1". In order to determine a new activation level, activation functions are applied to the net input. Functions that "squash" high values can be used to make sure that all new activations stay in the range of "0" to "1".
7. *Thresholds and biases*. Although activations can take on any value between "0" and "1", often thresholds and bias functions are used to force units to be either fully "on" or fully "off".
8. *A learning rule*. The basic goal of training is to bring the neural net into a state where it can take a given input and produce the correct output. To do

this, a learning rule is used to change the weights on the connection Supervised learning rules need to rely on the presence of a target output a the model for this changing of weights. Unsupervised learning rules do n rely on targets and correction, but use the structure of the input as the guide to learning.

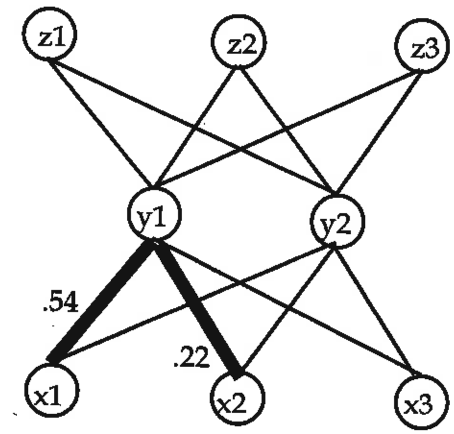

Figure 1. *A Sample Adaptive Neural Network.*

All connectionist networks share this common language of units, connections, weights, and learning rules. However, architectures differ markedly both in their detailed patterns of connectivity and in the specific rules used for activation and learning. For excellent, readable introductions to the theory and practice of neural network modeling, the reader may wish to consult Bechtel and Abrahamsen (1991) or Fausett (1994). For a mathematically more advanced treatment, see Hertz, Krogh, and Palmer (1991).

3. Local Lexical Maps

Nothing is more basic to language than the learning of new words. The child's first word often appears toward the beginning of the second year of life. But word learning is not a sudden process. Rather, it depends on a whole range of experiences and activities in which the child participates during the first year of life. Some of these experiences involve producing non-conventional sounds through babbling. Another type of experience involves listening to the cadences and phonetic forms of the words used by the adult community. Still another type of experience involves the slow development of thinking about the various categories of objects and events in the natural world. All of these activities and experiences are prerequisites to the learning of the first words. About two or three months before the first productive words are produced, we find some evidence that the child has begun to acquire a passive comprehension of a few of the most common words of the language. For example, the 14–month-old who has not yet produced the first word, may show an understanding of the word "dog" by turning to a picture of a dog, rather than a picture of a cat, when the word "dog" is uttered. It is difficult to measure the exact size of this comprehension vocabulary in the weeks preceding the first productive word, but it is perhaps no more than 20 words in size.

During this early period of auditory learning, the child starts to form associations between certain auditory patterns and particular meaningful interpretations. In older models of lexical learning, the process of associating a sound with a meaning involved the trivial formation of a single link. For example, in Morton's (1970) Logogen Model, the learning of a new word requires nothing more than the linking up of one already available pattern or cluster to another. The idea that auditory and semantic patterns form coherent clusters seems to reflect real facts about the infant's cognition. On the semantic level, one could argue (Mervis 1984) that the child's previous experience with dogs has served to promote the consolidation of the concept of a "dog". On the phonological level, it also appears that repeated exposure to the consistent pattern of "dog" also leads to the emergence of a consolidated phonological pattern.

The self-organizing feature map (SOFM) framework of Kohonen (1982) and Miikkulainen (1990) provides us with a way of characterizing these early processes of semantic and phonological consolidation. In the framework of SOFM models, word learning can be viewed as involving the development of maps in which individual patterns can be stored and retrieved reliably. Three types of local maps are involved in word learning: auditory maps, meaning maps, and articulatory maps. Each of these three maps uses the same learning algorithm. Figure 2 illustrates the activation of a particular node in an auditory map.

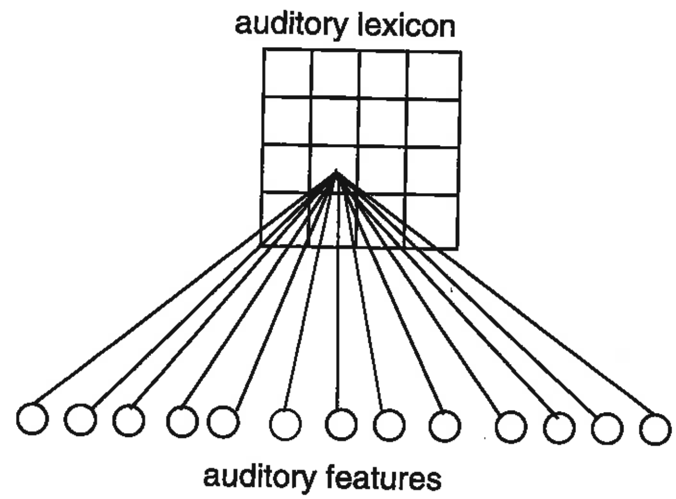

Figure 2. *A Self-organizing Feature Map for Storing Auditory Patterns.*

The input to this feature map involves a large number of auditory phonological features taken from separate domains such as sibilance, formant transition direction, formant duration, formant frequency, stop click timing, and others. These are schematically represented as "auditory features" at the bottom of Figure 2. For the purposes of computational modeling, the multidimensional space is compressed onto a 2–D topological space.

What makes this mapping process self-organizing is the fact that there is no pre-established pattern for these mappings and no preordained relation between particular nodes and particular feature patterns. The SOFM algorithm decides which node on the map should be the "winner" for a particular input pattern. At first, the weights on the map are set to small random values. When the first input comes in, the random setting of these weights makes it so that, by chance, some particular node is the one that is maximally responsive to the current input pattern. That node then decrements the activation levels on the other nodes. This decrementation takes on the form of a "Mexican hat" or sombrero. Right around the winner, related nodes are not decremented as much as are more distant nodes. Because of the architecture of the relation between the input and the grid, nodes that are nearby in the map come to respond to similar input patterns. For example, words that begin with similar initial segments will tend to be assigned to neighboring units in the map. The Mexican hat shape obeyed by the competitive interactions in the SOFM conforms closely to known facts about lateral inhibition and the redistribution of syntactic resources (Kohonen 1982) in cortical tissue. The actual computational implementation of this framework uses a computationally efficient algorithm that is faithful to these biological principles (Miikkulainen 1990).

This system works well to encode large numbers of patterns. In one sample simulation, we found that a 100 x 100 network with 10 000 nodes can learn up to 6 000 phonological patterns with an error rate of less than 1%. In this implementation, we used eight floating-point numbers to generate the input. At the beginning of learning, the first input vector of eight numbers led by chance to somewhat stronger activation on one of the 10 000 cells. This one slightly more active cell then inhibits the activation of its competitors, according to the Mexican hat function. As a result of this pattern of activation and inhibition, inputs that are close in feature space end up activating cells in similar regions of the map. Once a cell has won a particular competition, its activation is negatively dampened to prevent it from winning for all of the inputs. Then, on the next trial, another cell has a chance to win in the competition for the next sound-meaning input pattern. This process repeats until all 6 000 sound-meaning patterns have developed some "specialist" cell in the feature map. During this process, the dynamics of self-organization make it so that items with shared features end up in similar regions of the feature map.

We tracked the development of the feature map by computing the average radius of the individual items. After learning the first 700 words, the average radius of each word was 70 cells; after 3 000 words, the radius was 8; after 5 000 words the radius was 3; and after 6 000 words the radius was only 1.5 cells. Clearly, there is not much room for new lexical items in a feature map with 10 000 cells that has already learned 6 000 items. However, there is good reason to think that the enormous number of cells in the human brain makes it so that the size of the initial feature map is not an important limiting constraint on the learning of the lexicon by real children. We have found that there is no clear upper limit on the ability of the SOFM to acquire more items, when it is given a larger dimensionality.

3.1 *Using Maps for Retrieval*

In order to model additional aspects of lexical structure, the basic SOFM architecture must be supplemented by additional connections. Miikkulainen

(1990) did this by training reciprocal connections on two maps using Hebbian learning. Figure 3 illustrates the relations of these two maps. In this figure, a particular auditory form is associated with a particular semantic form or meaning.

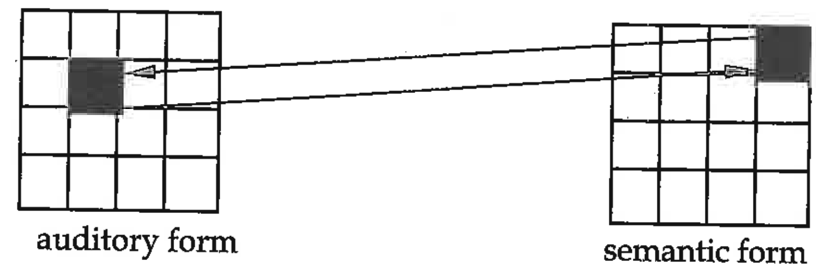

Figure 3. *A Bidirectional Sound-meaning Association in a Feature Map.*

Since neuronal connections can only fire in a single direction, training has to be conducted separately in each direction. In our simulations, learning begins with the consolidation of both the auditory and semantic maps according to the SOFM competitive learning algorithm. Once patterns are established on the two basic maps, Hebbian learning strengthens connections between units that are coactive on the sound map and the meaning map. This training is intended to represent the actual process of word learning, during which the child hears a word at the same time some meaningful aspect of the environment is being focused upon.

This proposed model is oversimplified in terms of both structure and process. In structural terms, additional maps are needed to represent additional aspects of lexical knowledge. In addition to the two maps given in Figure 3, there must be a map that encodes output phonological form, since the child must not only associate an auditory form to a semantic form, but must also associate the auditory form to an articulatory form and an articulatory form to the semantic form. Later, when the child learns to read and spell, there will also be maps for orthographic and visual forms. In processing terms, the SOFM given in Figure 3 fails to express important aspects of the serial structure of auditory and articulatory patterns. Later, we will discuss a lexical learning model developed by Gupta and MacWhinney (1997) that deals in a more explicit way with issues of serial ordering.

3.2 *Articulatory Scaffolding*

The relation between a pattern in the auditory map and a pattern in the semantic map is essentially arbitrary. There is nothing about the phonological shape of

/kæt/ that corresponds in some patterned way to the meaning "cat". However, the relation between auditory and articulatory forms is far more systematic. Once an adult has been exposed to a new auditory form, the corresponding articulatory form is extremely easy to produce. When we hear someone say that their last name is "Tomingo" we can quickly reproduce that name, even after only one trial.

For the child, the mapping from a new auditory form to an articulatory form is a bit more difficult, but it is still the case that audition serves to "scaffold" articulation. What this means is that the auditory form remains an active target as we attempt to match the form in articulation. By then listening to our articulation, we can verify the match of our output to the target auditory form. This allows us to correct errors and to set up an excitatory feedback loop between the two forms that stabilizes the new articulatory shape. Gupta and MacWhinney (1997) show how the development of this correspondence is based primarily on the mapping of correspondences between auditory fragments and articulatory fragments. In the simplest case, these fragments are syllables. For example, once the child has learned how to produce the syllable /go/ of "go", this auditory-articulatory correspondence is available for use in any new word. Even individual segments can be extracted through analysis. Some of this learning occurs during late babbling, but it is consolidated with the first words. Over time, the links between auditory and articulatory forms become more extensive.

3.3 *Prosody and Time*

Both the auditory and the articulatory maps must be structured to deal effectively with multisyllabic patterns. In order to process multisyllabic words, the input to the basic lexical map needs to derive from preprocessing by a SOFM which identifies individual syllables. This map stores a large number of identifiable syllabic forms such as /ba/, /kib/, and /Uv/, as well as subsyllabic forms such as /s/ or /n/. The input to this SOFM arrives in a sequential way, but each syllable is processed as a separate temporal chunk. This is easy to do on the level of the syllable, because there are many cues that tell whether a segment is in the position of the onset, the nucleus, or the coda. Because most coarticulation effects occur within the syllable, this is an effective way of dealing with low-level context effects. The syllabic processor operates repeatedly through the word to encode a series of activations of syllables.

The functioning of this syllabic map is supplemented by a process that associates particular syllabic vectors with additional prosodic information. This processor attends not to the segmental forms in the speech wave, but to the overall prosodic structure. Prosodic information works in terms of the system of metrical feet to encode the status of a given syllable as being in an iambic or trochaic foot and being either a strong or weak syllable. It is the union of these prosodic features with the basic segmental syllabic features which then serve as input to the auditory lexical SOFM. In a word like "banana" the syllabic processor operates repeatedly to encode three syllables. However, without the additional metrical information, these three encodings could be perceived as the patterns "nabana" or "nanaba", as well as "banana". In order to uniquely encode "banana", the first syllable /ba/ must be coded as a first foot weak beat, the second syllable /na/ must be coded as the strong beat and the final /na/ must be coded as the second foot weak beat. Thus, the complete input to the lexical map includes both segmental and prosodic information. It is this complete merged pattern which is then associated to the semantic pattern to specify emergent lexical items.

3.4 *Acquiring Inflectional Markers*

The local lexical map can be used to acquire not only stems such as "dog" or "jump" but also affixes such as the plural suffix or the past tense suffix. Stems can be learned directly. However, in order to model the learning of affixes, we need to examine an additional process called "masking" (Burgess 1995; Burgess & Hitch 1992; Carpenter, Grossberg & Reynolds 1991; Cohen & Grossberg 1987; Grossberg 1987). Let us use the learning of the English past tense suffix to illustrate how masking works.

1. The net learns a set of present tense verbs, along with the corresponding past tense forms. We can refer to this initial phase of learning as "rote" learning. These rote-learned forms include regular pairs such as "jump — jumped" and "want — wanted", as well as irregulars such as "run — ran" and "take — took".
2. The network then learns a new present tense such as "push" for which the corresponding past tense form has not yet been learned.
3. Then the child hears the word "pushed" with the auditory form /pUS/ and the semantic pattern "push + past". On the auditory map, the node corresponding to /pUS/ is the closest match. On the semantic map, the node corresponding to "push + present" is the closest match.
4. A pattern of bidirectional activation is established between the two maps. It is this bidirectional activation that supports the process of "masking".

Masking works to drain activation from nodes and features that are coactive in the two maps. In the current example, the features of the stem on both maps are all masked out, leaving the feature "past tense" as unmasked on the semantic map and the features corresponding to the final /id/ as unmasked on the auditory map.

5. The unmasked phonology is then associated with the unmasked semantics through the same type of Hebbian learning that is used to produce the basic rote-learning of new lexical forms.

This implements in a neuronally plausible way the process of morphological extraction by analysis. In the terms of MacWhinney (1978), affix analysis involves associating the "unexpressed" with the "uncomprehended". This approach to the problem of learning the English past tense solves two problems faced by earlier nonlexical models. First, the model succeeds in capturing both rote lexicalization and combinatorial lexicalization within a single connectionist model. Rote forms are picked up directly on the feature map. Combinatorial forms are created by the isolation of the suffix through masking and the use of masking in production.

Second, having learned to comprehend the past tense in a productive way, the child can then learn the association between the auditory pattern and an articulatory representation. This occurs when the child tries to produce the new form. The activation of a semantic pattern leads to the activation of an auditory pattern which then sets up a temporary excitatory feedback loop to the articulatory map. During the process of scaffolding, the auditory form remains active as we attempt to match the form in articulation. By then listening to our articulation, we can verify the match, correct errors, and set up an excitatory feedback loop between the two forms that stabilizes the new articulatory shape. As we noted earlier, the process of developing a match between the auditory and articulatory forms proceeds syllable by syllable by relying on prosody to encode the temporal properties of successive syllables.

3.5 *Inflectional Marking and the Logical Problem*

In the network we have been discussing, a single lexical feature map can produce both a rote form like "went" and a productive form like "*goed". The fact that both can be produced in the same lexical feature map allows us to begin work on a general solution to the "logical problem of language acquisition" (Baker & McCarthy 1981; Gleitman 1990; Gleitman, Newport & Gleitman 1984; Morgan & Travis 1989; Pinker 1984; Pinker 1989; Wexler & Culicover 1980). In the case of the competition between "went" and "*goed", we expect "went" to become solidified over time because of its repeated occurrence in the input. The form "*goed", on the other hand, is supported only by the presence of the -ed form. Figure 4 illustrates this competition:

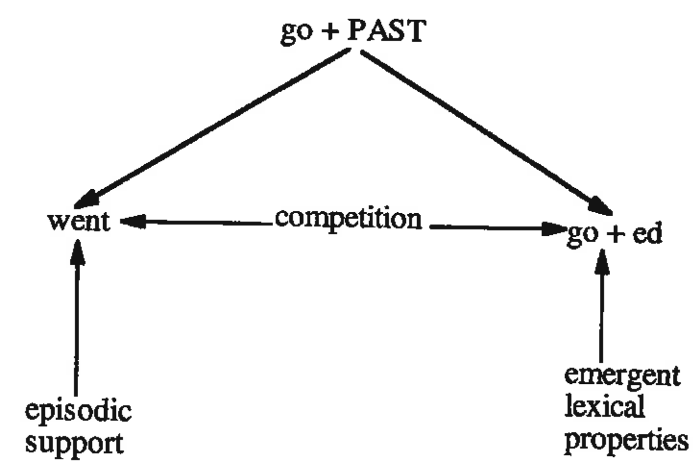

Figure 4. *Competition between Episodic and Combinatorial Knowledge.*

This particular competition is an example of what Baker and McCarthy (1981) calls a "benign exception to the logical problem". The exception is considered benign because the child can learn to block overgeneralization by assuming that there is basically only one way of saying "went". This Uniqueness Constraint is thought to distinguish benign and non-benign exceptions to the logical problem. However, from the viewpoint of the Competition Model account we are constructing here, all exceptions are benign.

The basic idea here is that, when a child overgeneralizes and produces "*goed", the system itself contains a mechanism that will eventually force recovery. In cases of overgeneralization, alternative expressions compete for the same meaning. One of these forms receives episodic support from the actual linguistic input. This episodic support grows slowly over time. The other form arises productively from the operation of analogistic pressures. When episodic support does not agree with these analogistic pressures, the episodic support eventually comes to dominate and the child recovers from the overgeneralization.

This is done without negative evidence, solely on the basis of positive support for the form receiving episodic confirmation.

4. Lexical Groups

The second level of linguistic structure we will discuss is the level of the lexical group. The formation of Level 2 lexical groups is an emergent process that depends on the existence of a Level 1 substrate of lexical items organized into SOFMs. The force that drives the emergence of lexical groups and their related syntactic properties is the linking of words into morphological and syntactic combinations. We can refer to the properties that emerge in this way as "emergent lexical properties". In this section, we will review some of these emergent properties.

4.1 *Inflectional Morphology and Lexical Groups*

Having acquired productive use of inflectional morphology, the child can begin to learn how to combine inflections with stems. The emergentist approach to language acquisition holds that the patterns governing these combinations emerge from information implicit in the lexical map. To illustrate how this works, let us take as an example the network model of German gender learning developed by MacWhinney, Leinbach, Taraban, and McDonald (1989). This network is designed to model how German children learn how to select one of the six different forms of the German definite article: "der", "die", "das", "des", "dem", or "den". Which of the six forms of the article should be used to modify a given noun in German depends on three additional features of the noun: its gender (masculine, feminine, or neuter), its number (singular or plural), and its role within the sentence (subject, possessor, direct object, prepositional object, or indirect object). To make matters worse, assignment of nouns to gender categories is often quite nonintuitive. For example, the word for "fork" is feminine, the word for "spoon" is masculine, and the word for "knife" is neuter.

Although these relations are indeed complex, MacWhinney *et al.* show that it is possible to construct a neural network that learns the German system from the available cues. The MacWhinney *et al.* model, like most current connectionist models, involves a level of input units, a level of hidden units, and a level of output units (Figure 5). Each of these levels or layers contains a number of discrete units or nodes. For example, in the MacWhinney *et al.* model, the 35

units within the input level represent features of the noun that is to be modified by the article. Each of the two hidden unit levels includes multiple units that represent combinations of these input-level features. The six output units represent the six forms of the German definite article.

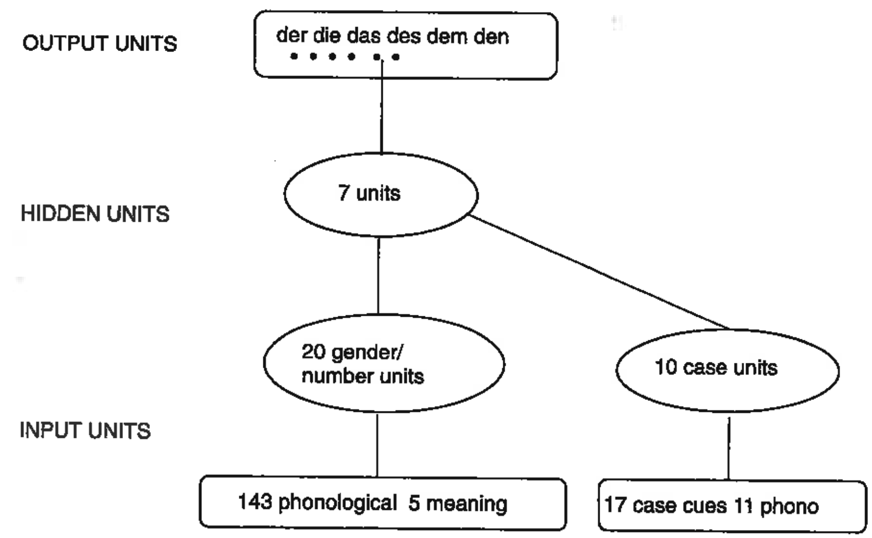

Figure 5. *A Back Propagation Model for German Declension.*

This network successfully learned its input corpus and displayed a good ability to generalize gender assignment to new nouns. It was also able to take a nominal form presented in one case and use it to predict a form in another case. The overgeneralization patterns the model produced matched up well with those produced by children. Despite its successes, this model and a similar model for English (MacWhinney & Leinbach 1991) faced certain basic problems. These problems all arose from the fact that the model assigned no privileged role to words as lexical items. Instead, all learning was based on an input composed of phonological patterns. A clear example of this type of problem arises in the case of the sound /rIN/ which represents three different verb meanings (to "ring" a bell, to "wring" out the clothes, and to "ring" the city with troops). The past tense forms of these verbs will be "rang", "wrung", and "ringed" depending on the meaning of the stem. By itself, the back propagation net cannot distinguish homophonic relations of this type. However, when a Level 2 back propagation network is joined to a Level 1 SOFM, homophony is no longer a problem, because the various homophonic meanings of "ring" are now representationally distinct in terms of the input that the SOFM can provide to the back propagation network. The reason for this is simply that the lexical SOFM contains semantic information which can be passed on to promote disambiguation in the back propagation network.

Gupta and MacWhinney (1992) showed how the addition of lexical information to the back propagation network for German leads to improved performance. Because the Gupta and MacWhinney model combines two different architectures, inflectional formations can be produced in several different ways. First, the SOFM can generate both regular and irregular forms by rote. Second, because the SOFM includes affixes along with stems, regular affixation can be produced through combination. Third, the pattern generalization processes found in the back propagation network can help produce irregularizations. For example, the past tense forms "wrung" and "rang" could be produced either directly by rote or by generalization using the back propagation network.

4.2 *Argument Frame Induction from Lexical Groups*

The strategy of linking a Level 2 back propagation network to the Level 1 lexical SOFM also helps us account for the learning of syntactic patterns. The Competition Model (MacWhinney 1988) has consistently emphasized the role of lexical argument ("valency" or "dependency") relations as the basic controllers of syntactic structure. This analysis was grounded originally on the theories of Lexical Functional Grammar (LFG) (Bresnan 1982) and Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994) that developed during the early 1980s. The role of lexical predicates in determining syntactic structure is now widely accepted. However, there is still no agreement regarding the ways in which children learn to attach argument frames to lexical items or groups of lexical items. Non-connectionist proposals regarding this learning can be found in Brent (1994), MacWhinney (1988), and Pinker (1984). Within a connectionist framework, the major attempts to deal with syntactic processing include Elman (1990), McClelland and Kawamoto (1986), Miikkulainen (1993), and St. John (1992). However, none of these accounts comes to grips with the relation between argument frames and specific lexical items.

We know that the induction of argument relations must occur in parallel with the process of learning new words. To illustrate this process, consider an example in which the child already knows the words "Mommy" and "Daddy", but does not know the word "like". Given this state of lexical knowledge, the sentence "Daddy likes Mommy" would be represented in this way:

d a d i		l aɪ k s		m a m i
Daddy	|	unknown	|	Mommy

For the first and third phonological stretches, there are lexical items that match. These strings and the semantics they represent are masked. The unknown stretch is not masked and therefore stimulates lexical learning of the new word "likes". The core of the learning for "likes" is the association of the sound /laɪk/ with the meaning "like". In addition to this basic Level 1 lexical association, the child must also construct additional links to Level 2 argument relations. At first these patterns are grounded on a few lexical items. However, these Level 2 patterns quickly generalize to apply to lexical groups. The initial lexical argument frame for the word "likes" is:

arg1: preposed, "Daddy", experiencer
arg2: postposed, "Mommy", experience

Further exposures to sentences such as "Daddy likes pancakes" or "Billy likes turtles" will soon generalize the dependency frame for "likes" to:

arg1: preposed, experiencer
arg2: postposed, experience

No theoretical weight is placed on the notion of "experiencer" or "experience" and different learners may conceptualize this role in different ways.

Adjectives typically have only one argument. Prepositions have two — one for the object of the preposition and a second for the head of the prepositional phrase. Verbs can have as many as three arguments. For each lexical item, we can refer to these arguments as arg1, arg2, and arg3. When a group of words share a common set of semantic relations with a particular argument, they form a lexical group argument frame, or, more succinctly, a "group frame". For example, words like "send" or "promise" share the syntactic property of permitting a double object construction as in "Tim promised Mary the book". Pinker (1989) and others have argued that there are a variety of semantic cues which work together to decide which verbs allow this type of double object construction.

4.3 *Relations between Level 1 and Level 2*

Level 1 information is stored in SOFMs and Level 2 information is organized into back propagation networks dependent on Level 1 information. Figure 6 sandpaper to something. MacWhinney (1989) talks about these semantic extension effects in terms of a process of "pushy polysemy".

5. Functional Neural Circuits

The third level of neurolinguistic structure is the level of the functional neural circuit. This level requires the integration of information across large distances in the cerebral cortex. A prototypical example of a functional neural circuit is the phonological rehearsal loop that supports verbal short-term memory (Gathercole & Baddeley 1993; Gupta & MacWhinney 1994). Recent work with neural imaging (Grasby *et al.* 1993; Paulesu, Frith & Frackowiak 1993) indicates that this loop is based on the coparticipation of auditory processing areas in the superior temporal gyrus, attentional regions in the frontal cortex, and articulatory areas in the motor cortex. Similar posterior-frontal functional neural circuits have also been identified in visual processing.

Unlike Level 1 and Level 2 processing, the type of processing that requires the use of functional neural circuits can place severe demands on attentional resources. As long as sentence processing can emerge from Level 2 use of argument frame structures, a minimal demand is placed on additional attentional resources. As each predicate is linked to its several arguments, the listener shifts focus away from the individual lexical items onto the emerging sentence interpretation (Gernsbacher 1990; MacWhinney 1977). In effect, every word that is linked to the growing interpretation is "masked" in Level 1 lexical maps. This type of local processing is highly automatic and essentially effortless. However, some syntactic structures place a heavy demand on working memory. For example, in a sentence such as "The dog the cow the pig chased kicked barked", the listener cannot construct interpretations by linking each word to its neighbor. Instead the string of three nouns and three verbs have to be stored in unassociated form in working memory, while the listener attempts to find meaningful clusters. Sentences of this type, while technically grammatical, are notoriously difficult to process. Accumulations of unattached nouns in relative clauses are a well-known problem for speakers of SOV languages such as Hungarian (Mac Whinney & Pléh 1988) and Japanese (Hakuta 1981).

5.1 *Conservatism, Functional Circuits, and the Logical Problem*

The Competition Model emphasizes the extent to which lexical competition can solve the logical problem of language acquisition. However, there are certain complex syntactic structures for which the lexical solution is more questionable. For example, O'Grady (1987) notes that children learn positive contexts for *wh*-movement in this order:

(1) What did the little girl hit __ with the block today?
(2) What did the boy play with __ behind his mother?
(3) What did the boy read a story about __ this morning?

Although one might be able to formulate a lexical basis for the processing of these *wh*-movement patterns, it is more likely that they involve a form of sentence memory that relies rather more on functional neural circuits and less on lexically-organized information. What is interesting is the fact that, precisely in these non-lexical contexts, children's tendency toward conservatism seems to be maximized. Children are never presented with contexts such as (4):

(4) *What did the boy with ___ read a story this morning?

Because children approach the learning of these contexts conservatively, they seldom make overgeneralizations of this type and seldom attempt *wh*-movement in this particular context. The general principle seems to be that overgeneralization occurs primarily with Level 2 argument frame patterns and not with Level 3 long-distance movement patterns. For Level 3 patterns, the attentional and computational difficulties involved lead children to adopt a conservative approach that minimizes the role of overgeneralization. This is not to say that overgeneralization of long-distance movement never occurs. However, numerically speaking, it is much rarer than argument frame overgeneralization. Because of this conservativism, attribution of language acquisition to innate knowledge of conditions blocking subjacency violations seems unmotivated.

Summary

At this point, it may be useful to summarize the core assumptions being made in this account of language emergence:

1. The model assumes an auditory processing mechanism that can extract information regarding the onset, nucleus, and coda elements of individual syllables.
2. The information from the syllabic processor is supplemented by information from the prosodic processor which marks the position of each syllable in terms of feet and beats.
3. Auditory and semantic information about words is encoded in a self-organizing feature map.
4. Associations between sound and meaning are formed through Hebbian learning.
5. Auditory information can be used to scaffold the construction of an articulatory representation. This is done in terms of syllables and prosodic structures.
6. Masking in lexical recognition provides the support for the extraction of new affixes.
7. Changes in stems and affixes can be controlled through a system of modifications using the back propagation algorithm.
8. Sentence interpretation requires the linking of words in terms of argument structures. These structures are learned through frame generalization in back propagation networks which receive input from the lexical map.
9. The processing of complex syntactic structures and lists of words requires the involvement of functional neural circuits including frontal attentional processing and temporal lobe verbal memory and rehearsal.

In this model of language development, the first commitment that the brain makes is to the encoding of auditory, articulatory, and lexical information in localized maps. After this information is consolidated, back propagation systems develop to fine-tune the interactions of lexical items, and functional neural circuits control capacity-intensive aspects of sentence processing.

Although the developments we have discussed lead to a great complexity of patterns and constructions, the underlying elements of feature maps, masking, argument frames, and rehearsal loops from which these patterns emerge are themselves cognitively basic structures grounded in fundamental properties of neural structure and functioning. Some aspects of these structures are probably basic to all of mammalian cognition. However, the great elaboration of lexical structures that we find in human language point to the extensive elaboration of earlier structures during the million years of human evolution. Most recently, the overlay of functional neural circuits between areas such as the frontal attentional

areas and the temporal auditory areas has led to further species-specific advances in the capacity for learning and using language. Moreover, the specific elaboration of lexical feature maps also appears to be a specifically human adaptation. Although this model tends to emphasize the cognitive adaptations involved in supporting language processing, it would be a mistake to ignore the important changes in social structure and interpersonal subjectivity that have also supported the evolution of human language. Hopefully, continuing rapid advances in our understanding of brain function and structure will allow us to soon begin to understand how these emotional and social underpinnings support the computational and cognitive structures we have discussed here.

These biological and cognitive aspects of an emergentist account of human language will eventually need to be related to the equally important social and discourse pressures that control the shape of grammar and the lexicon. Together, these various emergentist visions allow us to construct a new view of human language that goes beyond the simple debate between functionalism and formalism and emphasizes the interplay of alternative streams, mechanisms, and processes of emergence.

References

Baker, C. L. & J. J. McCarthy (Eds.). 1981. *The logical problem of language acquisition.* Cambridge: MIT Press.

Bechtel, W. & A. Abrahamsen. 1991. *Connectionism and the mind: An introduction to parallel processing in networks.* Cambridge, Mass.: Basil Blackwell.

Bowerman, M. 1988. "The 'no negative evidence' problem". In J. Hawkins (Ed.), *Explaining language universals,* (73–104). London: Blackwell.

Brent, M. 1994. "Surface cues and robust inference as a basis for the early acquisition of subcategorization frames". *Lingua 92,* 433–470.

Bresnan, J. (Ed.). 1982. *The mental representation of grammatical relations.* Cambridge, Mass.: The MIT Press.

Burgess, N. 1995. "A solvable connectionist model of immediate recall of ordered lists". In G. Tesauro, D. Touretzky & J. Alspector (Eds.), *Neural Information Processing Systems 7,* (1–7). San Mateo, CA: Morgan Kaufmann.

Burgess, N. & G. Hitch. 1992. "Toward a network model of the articulatory loop". *Journal of Memory and Language 31,* 429–460.

Carpenter, G., S. Grossberg & J. Reynolds. 1991. "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network". *Neural Networks 4*, 565–588.

Cohen, M. & S. Grossberg. 1987. "Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data". *Applied Optics 26*, 1866–1891.

Du Bois, J. 1987. "The discourse basis of ergativity". *Language 63*, 805–856.

Elman, J. 1990. "Finding structure in time". *Cognitive Science 14*, 179–212.

Fausett, L. 1994. *Fundamentals of Neural Networks*. Englewood Cliffs, N.J.: Prentice Hall.

Gathercole, V. & A. Baddeley. 1993. *Working memory and language*. Hillsdale, N.J.: Lawrence Erlbaum.

Gernsbacher, M. A. 1990. *Language comprehension as structure building*. Hillsdale, N.J.: Lawrence Erlbaum.

Gleitman, L. 1990. "The structural sources of verb meanings". *Language Acquisition 1*, 3–55.

Gleitman, L. R., E. L. Newport & H. Gleitman. 1984. "The current status of the motherese hypothesis". *Journal of Child Language 11*, 43–79.

Grasby, P. M., C. D. Frith, K. J. Friston, C. Bench, R. S. J. Frackowiak & R. J. Dolan. 1993. "Functional mapping of brain areas implicated in auditory-verbal memory function". *Brain 116*, 1–20.

Grossberg, S. 1987. "Competitive learning: From interactive activation to adaptive resonance". *Cognitive Science 11*, 23–63.

Gupta, P. & B. MacWhinney. 1992. Integrating category acquisition with inflectional marking: A model of the German nominal system, *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, N.J.: Lawrence Erlbaum.

Gupta, P. & B. MacWhinney. 1994. "Is the articulatory loop articulatory or auditory? Re-examining the effects of concurrent articulation on immediate serial recall". *Journal of Memory and Language 33*, 63–88.

Gupta, P. & B. MacWhinney. 1997. "Vocabulary acquisition and verbal short-term memory: Computational and neural bases". *Brain and Language 59*, 267–333.

Hakuta, K. 1981. "Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese". *Cognition 9*, 197–236.

Hertz, J., A. Krogh & R. Palmer. 1991. *Introduction to the theory of neural computation*. New York: Addison-Wesley.

Hopper, P. J. & S. A. Thompson. 1984. "The discourse basis for lexical categories in universal grammar". *Language 60*, 703–752.

Kohonen, T. 1982. "Self-organized formation of topologically correct feature maps". *Biological Cybernetics 43*, 59–69.

MacWhinney, B. 1978. "The acquisition of morphophonology". *Monographs of the Society for Research in Child Development 43*, Whole no. 1, pp. 1–123.

MacWhinney, B. 1988. "Competition and teachability". In R. Schiefelbusch & M. Rice (Eds.), *The teachability of language*, (63–104). New York: Cambridge University Press.

MacWhinney, B. 1977. "Starting points". *Language 53*, 152–168.

MacWhinney, B. 1989. "Competition and lexical categorization". In R. Corrigan, F. Eckman & M. Noonan (Eds.), *Linguistic categorization*, (195–242). New York: Benjamins.

MacWhinney, B. J., J. Leinbach, R. Taraban & J. L. McDonald. 1989. "Language learning: Cues or rules?". *Journal of Memory and Language 28*, 255–277.

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

MacWhinney, B. & C. Pléh. 1988. "The processing of restrictive relative clauses in Hungarian". *Cognition 29*, 95–141.

McClelland, J. L. & A. Kawamoto. 1986. "Mechanisms of sentence processing: Assigning roles to constituents". In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing*, (272–376). Cambridge, Mass.: MIT Press.

Mervis, C. 1984. "Early lexical development: The contributions of mother and child". In C. Sophian (Ed.), *Origins of cognitive skills*, (339–370). Hillsdale, N.J.: Lawrence Erlbaum.

Miikkulainen, R. 1990. A distributed feature map model of the lexicon, *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale, N.J.: Lawrence Erlbaum.

Miikkulainen, R. 1993. *Subsymbolic natural language processing*. Cambridge, Mass.: MIT Press.

Morgan, J. & L. Travis. 1989. "Limits on negative information in language input". *Journal of Child Language 16*, 531–552.

Morton, J. 1970. "A functional model for memory". In D. A. Norman (Ed.), *Models of Human Memory*, (203–248). New York: Academic Press.

O'Grady, W. 1987. *Principles of grammar and learning*. Chicago: Chicago University Press.

Paulesu, E., C. D. Frith & R. S. J. Frackowiak. 1993. "The neural correlates of the verbal component of working memory". *Nature 362*, 342–345.

Pinker, S. 1984. *Language learnability and language development*. Cambridge, Mass: Harvard University Press.

Pinker, S. 1989. *Learnability and cognition: the acquisition of argument structure*. Cambridge: MIT Press.

Pollard, C. & I. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: Chicago University Press.

St. John, M. 1992. "The story gestalt: a model of knowledge-intensive processes in text comprehension". *Cognitive Science 16*, 271–306.

Wexler, K. & P. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, Mass. : MIT Press.

Underspecification and Modularity in Early Syntax

A formalist perspective on language acquisition

Nina Hyams
UCLA

Abstract

In this paper, we will review a range of cross-linguistic empirical evidence supporting the view that "telegraphic" children have a rich and complex syntax, including knowledge of functional structure and of language-specific parameter values associated with functional categories. We argue that the optionality of functional elements — finiteness, determiners, subject pronouns — in early language arises through the interaction of a rather well-developed grammar with an immature pragmatic system. We will show that the expression of certain functional elements is dependent on the expression of others, for example, finiteness on the verb is contingent upon the type of subject. We suggest that functionalist accounts of early language, as well as performance accounts which attribute the omission of these elements to a processing bottleneck, fail on empirical grounds. The syntactic regularities which are observed are most adequately explained within a modular framework. We will also briefly discuss the *logical problem of language acquisition* (LPLA), in connection with functional underspecification.

1. Introduction: Modularity and language development

One of the most striking aspects of early language is the apparent optionality of various functional elements, such as pronouns, verbal inflection, and determiners. These elements, and the syntactic categories that contain them, constitute the syntactic frame of the sentence, providing a skeleton for the "meatier", more