The Emergence of Language from Body, Brain, and Society

Brian MacWhinney

1 Stipulationism

Scientists respect accurate descriptions. For example, geologists appreciate it when a scientific report describes the range of variation in the sediments and fossils of the Florissant sandstones in the Rocky Mountains in sufficient detail to allow them to quickly and accurately judge whether a new outcrop of sandstone comes from the Florrisant or some other formation of the Mesozoic. However, no geologist would confuse an inventory of fossils in a particular rock bed with an explanatory theory of orogeny or plate tectonics. Unfortunately, workers in cognitive science have found it more difficult to distinguish description from explanation. Throughout the second half of the last century, workers in generative grammar and artificial intelligence constructed complex rule-based models of mental processing that used the stipulation of rules, processes, and constraints as a way of achieving what was thought to be an explanation of psychological process. On a descriptive and practical level, this work was often successful. However, with few exceptions {Newell, 1990 #5300}, rule-based systems failed to provide fundamental explanations of cognitive phenomena.

In the sciences, explanation of physical and biological structures typically rests upon an elucidation of the underlying forces and dynamics from which these structures emerge. For example, the activities of the volcanoes rimming the Pacific are explained by theories of crustal subduction in the context of the larger theory of plate tectonics. The shape of crystalline structures such as pyrite, diamond, or salt is explained by the geometry of the electronic bonds formed by their component molecules, as well as by the physical size of the molecules that are packed into the crystal lattice. From these basic physical constraints, there emerges a set of eight typically crystalline habits, each with a set of possible deformations. To provide descriptive detail to these explanatory accounts, crystallographers and other scientists use mathematical models that express the imposition of the underlying constraints.

Within linguistics, physicalist explanation has played a major role at the interface between experimental phonetics and phonology {Ohala, 1974 #3165}. In that field, the shape of tones, syllables, and assimilations is viewed as emerging from the interaction of various physical constraints on sound production. Other branches of linguistics have explicitly avoided grounding linguistic structure on external constraints. Fields like generative syntax, generative phonology, and Montague grammar have claimed that the explanation of structure within a particular language module can be provided by examining constraints arising from the structure of the module itself. From the emergentist point of view, this would be much like arguing that the form and activities of particular volcanoes is derivable from autonomous principles of volcano composition or that the shape of crystals is derivable from autonomous principles of crystal form. Crystallographers would not object to a summary of the various crystal habits as a description, but they would not confuse this descriptive summary with a fuller explanation based on bond patterns and molecular size. In all such cases, the structures being described are viewed as emerging from the imposition of external constraints.

In the 1990s, researchers began to explore ways of applying this same emergentist framework to language structure. The first attempts in this direction, attempted to ground the processing of linguistic rules on models of neural networks {Rumelhart, 1987 #3563}. These neural network models viewed children as learning cues, rather than rules {MacWhinney, 1989 #2727}. Soon, researchers began to explore still other emergentist alternatives to rule systems, including dynamics systems theory {Thelen, 1994 #6981}, data-driven learning {Bybee, 2001 #9518}, and biological models of neural plasticity {Elman, 1999 #8623}. At the same time, linguistic theory also began to move away from stipulationism, attempting to extract a minimal set of principles from which broader syntactic patterns could emerge {Chomsky, 1995 #7306;Kager, 1999 #9063}.

2 Emergence from what?

The examples we have considered so far all involve physical structures that emerge from physical constraints. However, emergence is not limited to this type of relation. Let us consider several quite different forms of emergence.

First, consider a case of the emergence of a physical pattern from social behavior. If you spend time watching the checkout lines at a supermarket, you will find that the number of people queued up in each line stays roughly the same. There are rarely six people in one line and two in the next. There is no socially articulated rule governing this pattern. Instead, the uniformity of this simple social "structure" emerges from other basic facts about the goals and behavior of shoppers and supermarket managers.

Honeybees are certainly no smarter than shoppers. However, working together, bees are able to construct an even more complex structure. When a bee returns to the hive after collecting pollen, she deposits a drop of wax-coated honey. Each of these honey balls is round with approximately the same size. As these balls get packed together, they take on the familiar hexagonal shape that we see in the honeycomb. There is no gene in the bee that codes for hexagonality in the honeycomb, nor any overt communication regarding the shaping of the cells of the honeycomb. Rather, this hexagonal form is an emergent consequence of the application of packing rules to a collection of honey balls of roughly the same size.

Biological patterns emerge in very similar ways. For example, the shapes of the spots on a leopard or the stripes on a tiger emerge from the timing of the expression of a pair of competing genes expressing color as they operate across the developing leopard or tiger embryo {Murray, 1988 #9040}. No single gene directly controls these patterns. Rather, the stripes emerge from the interactions of the gene products on the physical surface of the embryo. Similar constraints operate during neurogenesis as cells migrate from the germinal matrix to their final locations in the cortex, leaving axonal projections along their way that form the white matter of the brain.

Once physical constraints and gene product interaction determine the basic connectivity pattern of the brain, the fine structure is shaped by further local physical competition. For example, Miller, Keller, and Stryker {, 1989 #5066} have shown how the ocular dominance columns described by Hubel and Weisel {, 1963 #7114} in their Nobel-prize-winning work can emerge as a solution to a competition between projections from the different optic areas during synaptogenesis in striate cortex. Further into development, physicalist models from Elman {, 1999 #8623} and Li, Farkas, and MacWhinney {Li, 2003 #9590} show how the learning of linguistic categories can emerge from the interaction of constraints on the activation of sheets of neuronal tissue. As development continues, Quartz and Sejnowski {, 1997 #7200} have shown local forces can lead to synaptic sprouting that provides for ongoing neuronal plasticity, and recent fMRI work {Booth, 1999 #8994} has shown how children with early brain lesions use a variety of alternative neuronal pathways to preserve language functioning.

3 Generativity

At this point, one might well object that language is fundamentally different from honeycombs and crystals. For these structures there is a finite set of meaningful structural variations. Language, on the other hand, has an infinite structure. The core teaching of generative grammar is that the infinite productivity of generative grammar arises from the operation of recursion in an autonomous syntactic module. Emergentism does not challenge the importance of recursion in producing infinite productivity. However, rather than simply attributing recursion to the remarkable functioning of an autonomous syntactic model, emergentism seeks to provide a physicalist account of the evolution of recursion {MacWhinney, 2003 #9477}. To develop this account, we must first determine the ways in which the brain provides the working memory components that must underlie recursion. An initial account of this type is developed in Gupta and MacWhinney {, 1997 #6908}. However, by itself working memory is not enough to support a fully grounded account of recursion. It is also important to understand how recursion emerges from the shifting of perspectives during sentence processing {MacWhinney, 1999 #7785}. Generativity is then viewed as emerging from recursion, which in turns emerges from the processes of working memory and perspective taking, which in turn emerge from the neural structures that subserved earlier nonlinguistic cognition.

4 Caveats

As we begin to explore the development of emergentist accounts for language structure, we must temper our enthusiasm in three ways. First, we must be careful not to think that any model of language structure or development that does not stipulate particular rules or hard-wired modules is emergentist. If we simply used the absence of stipulated rules as our criterion, we would allow even the most half-baked, inarticulate idea to count as an emergentist solution. Crucially, true emergentist accounts must specify particular mechanisms from which a structure develops. Ideally, we must be able to run a computational or numerical simulation of this mechanism that produces the observed patterns in the data. However this requirement leads to the second problem with emergentist accounts. This is the fact that, in many cases, they are exceedingly complex. For example, a state-of-the-art physicalist simulation of the functioning of the vocal tract requires the coordination of some 60 differential equations modeling the actions of the glottis, the pharynx, the tongue, and the other vocal structures. Of course, this type of problem is not unique to models of language functioning. Programs for simulating seismic activity and weather patterns are similarly complex and computationally difficult.

In some areas, the formation of emergentist accounts is limited not just by computational complexity, but also by the lack of relevant descriptive information. Even for a well-studied and well-mapped neurological structure such as the hippocampus, we are still unsure how the individual local areas interact to produce effective processing of memories. Given this uncertainty about the function of basic memory processes, it is clear that accounts of the emergence of language structure from neuronal processing must be tentative and incomplete. However, our uncertainty regarding particular mechanisms should lead us to give up on the search for explanation.

The final caveat we need to mention regards the issue of the falsifiability of emergentism. The core claim here is that all linguistic structures arise from the dynamic operation of external pressures from the body, the brain, and society. I do not believe that this general conceptualization can be falsified any more than the basic scientific method itself can be disproven. However, it is certainly possible to falsify claims regarding the emergence of specific structures from specific mechanisms. This then becomes the focus of linguistic work for the 21st century.

5 Mechanisms

As we move to replace the earlier stipulationism with the new emergentism, we need to focus on developing a fuller understanding of the arsenal of basic emergent mechanisms. In the end, all emergentist accounts must be grounded on these core mechanisms. Some examples of general mechanisms include:

1. Learning through error propagation. A good example of this type of mechanism is the back-propagation algorithm used in PDP modeling {Rumelhart, 1986 #2853}.
2. Self-organization. Mechanisms such as the self-organizing feature map {Kohonen, 1990 #7231} provide alternatives to mechanisms based on error propagation.
3. Item-based learning. In the area of grammatical development, the theory of item-based learning {MacWhinney, 1975 #2683;Tomasello, 2000 #9481} relies on general concepts from Construction Grammar {Goldberg, 1999 #8629}.
4. Reorganization of cognitive function to the contralateral hemisphere. Children with early left focal lesions are able to recover language function by reorganizing language to the right hemisphere. This plasticity in development is a general mechanism that supports a wide variety of emergent responses to injury or sensory disability {Booth, 1999 #8994;MacWhinney, 2000 #7795;Corina, 1992 #6073}.
5. Physical pressures on cognitive structures. Phonologists have shown that the shape of the vocal mechanism has a wide-ranging impact on phonological processes {Ohala, 1974 #3165}. Rather than stipulating phonological rules or constraints {Kager, 1999 #9063;Bernhardt, 1998 #9054}, we can view them as emergent responses to these underlying pressures.
6. Conversational emergence. Linguistic structures seem to be adapted to specific conversational patterns as they emerge online. For example, Du Bois {, 1987 #1135}

has argued that ergative marking in languages emerges from the fact that speakers tend to delete the actor in transitive sentences, because it is already given or known.

7. Perceptual recording. Recent studies of infant auditory perception {Jusczyk, 1997 #9056} have revealed that, even in the first few months, infants apply some general-purpose mechanism to record and learn from auditory input.
8. Constituent structure. All syntactic theories need to assume that related words cluster together in units and that the head of those units then serves to cluster with higher argument slots. This fundamental process of constituent structuring must be based on a set of basic mechanisms for motor control and planning {Donald, 1999 #9467}.

This is, of course, just a small sampling of the many mechanisms and pressures that shape the emergence of language. Others involve the shape of social relations in the young child's family {Ninio, 1988 #3131}, the shape of the input to guest workers learning a second language {Klein, 1989 #4930}, the preference in the brain for short connections {Shrager, 1995 #7422}, and the shape of sound dissipation for low frequencies across distances. In each case, the mechanisms we are considering are either corroborated through direct observation or are highly general processes based on lower-level mechanisms that have been directly observed.

6 Domain Generality

Within the language learning community, there is an active debate regarding the extent to which language learning is based on domain-general mechanisms. Sabbagh and Gelman {, 2000 #9482} present an analysis which equates emergentism with domain generality. This strong formulation of the emergentist position matches up well with the disembodied connectionism of the 1980s {Rumelhart, 1986 #2853}. However, the strong version fails to fully appreciate the degree to which emergentists view cognition as grounded on the body, the brain, and the social situation.

Consider a simple example from phonological development. There is a universal tendency to avoid sequences of nasal consonants followed by voiceless obstruents, as might arise in forms like ‘manpower.’ This constraint is grounded on the facts of speech production {Huffman, 1993 #9065} and figures prominently in recent elaborations of Optimality Theory {Kager, 1999 #9063}. Languages use at least five phonological processes to deal with this problem. These processes include nasal substitution, post-nasal voicing, denasalization, nasal deletion, and vowel epenthesis. Initially, children may apply a variety of these processes {Bernhardt, 1998 #9054}. Which processes are preserved and which are dropped out will depend on the shape of the target language, be it Indonesian, Quechua, Toba Batak, English, or Kelantan Malay. The shape of the vocal tract and the innervation of the muscles of the tongue determine the domain-specific landscape. Domain-general processes sample these constraints and negotiate between them in real time. In the terms used by Sabbagh and Gelman, the overall system of constraint satisfaction is a ‘buzzsaw’ cutting patterns through the local domain of embodied articulatory constraints. This example emphasizes the extent to which emergentism must make reference to the body. To attempt to construct an emergentist psycholinguistics that ignores the body, the brain, and the social situation would be like attempting to build an emergentist account of honeycomb formation that ignores the honey.

Although it is clear that emergentism needs to refer to domain-specific facts about the body, it is not clear that it needs to rely on any domain-specific cognitive mechanisms. Instead, it is likely that evolution reuses general cognitive mechanisms to serve new functions in special areas. For example, Givón {, 1998 #9410} has argued that the major cognitive event that occurred during language evolution involved a linkage of episodic memory to the auditory system through the support or tutelage of the visual system. The visual system had already established general mechanisms for the episodic encoding of spatial position and form. Primates had already developed a mechanism for recording auditory sequences {Hauser, 2001 #9444}. Adapting this mechanism to the task of language learning involved reshaping and relinking previously available cognitive mechanisms. It is true that these domain general episodic mechanisms have a specific localized shape for each modality. However, it is likely that the general mechanisms undergo a special tuning when they function at the local level {Caplan, 1999 #9484;Pinker, 1999 #9483}.

7 Emergence in Grammar

In the next three sections, I will present three specific example emergentist solutions to central problems in language learning. These sections will examine, respectively, grammatical learning, lexical learning, and language evolution. In this section, we will look at how emergentism provides accounts for grammatical learning.

One of the most active areas in recent work on language acquisition has been the study of the child's learning of inflectional marking. In English, inflections are short suffixes that occur at the ends of words. For example, the word "dogs" has a final /s/ suffix that marks the fact that it is plural. There are now well over 30 empirical studies and simulations investigating the learning of inflectional marking. The majority of work on this topic has examined the learning of English verb morphology with a particular focus on the English past tense. These models are designed to learn irregular forms such as "went" or "fell", as well as regular past tense forms such as "wanted" and "jumped". Other areas of current interest include German noun declension, Dutch stress placement, and German participle formation. Although the learning of inflectional markings is a relatively minor aspect of language learning, our ability to quantify this process has made it an important testing ground not only for the study of child language, but for developmental psychology and cognitive science more generally.

To illustrate how connectionist networks can be used to study the learning of inflectional morphology, let us take as an example the model of German gender learning developed by MacWhinney, Leinbach, Taraban, and McDonald {, 1989 #2727}. This model was designed to explain how German children learn to select one of the six different forms of the German definite article. In English we have a single word “the” that serves as the definite article. In German, the article can take the form “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. Acquiring this system of arbitrary gender assignments is particularly difficult for adult second language learners. Mark Twain expressed his consternation at this aspect of German in a treatise entitled “The aweful German language” {Twain, 1935 #5589} in which he accuses the language of unfairness and capriciousness in its treatment of young girls as neuter, the sun as feminine, and the moon as masculine. Along a similar vein, Maratsos and Chalkley {, 1980 #2770} argued that, because neither semantic nor phonological cues can predict which article accompanies a given noun in German, children could not learn the language by relying on simple surface cues.

Although these relations are indeed complex, MacWhinney et al. show that it is possible to construct a connectionist 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 1). Each of these levels or layers contains a number of discrete units or nodes. For example, in the MacWhinney et al. model, the 176 units within the input level represent features of the noun that is to be modified by the article. The phonological units code the sound of the stem using a system of features in syllabic slots. The meaning units represent semantic features such as inherent masculinity for male animals. The case cues code the surface level features that determine the thematic role of the noun phrase, and the additional 11 phonological marking there are for the genitive and dative suffixes of German. 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 article.

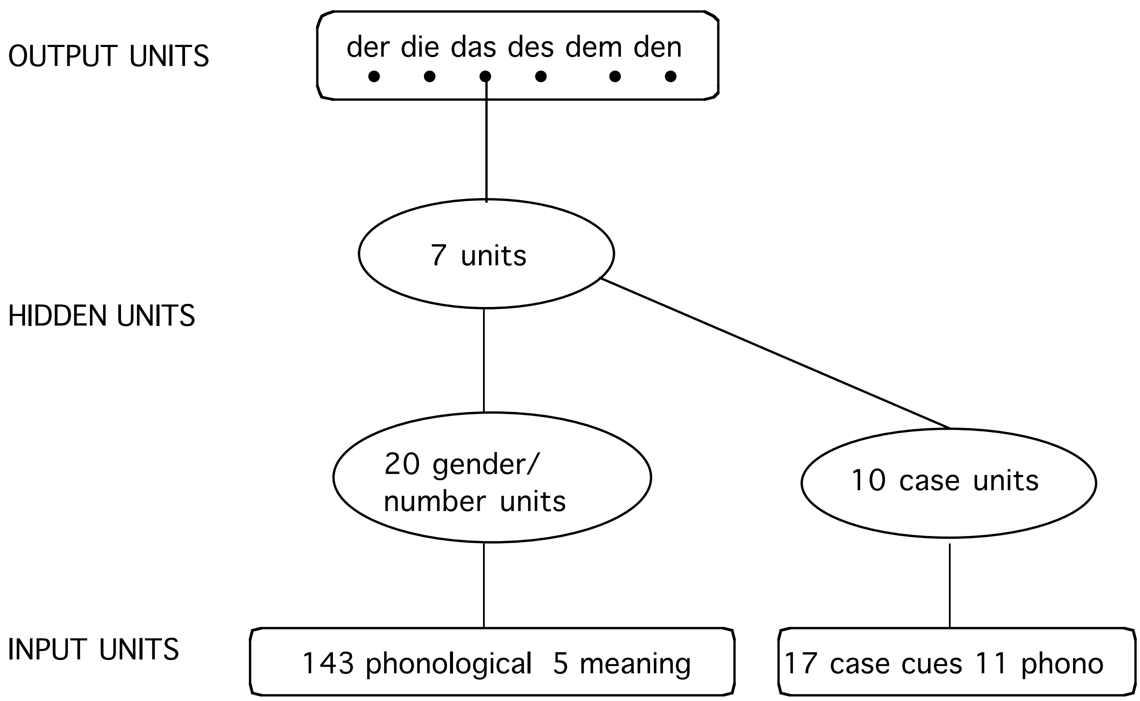

Figure 1: A network model of the acquisition of German declensional marking

As noted above, a central feature of such connectionist models is the very large number of connections among processing units. As shown in Figure 1, each input-level unit is connected to first-level hidden units; each first-level hidden unit is connected to second-level hidden units; and each second-level hidden unit is connected to each of the six output units. None of these hundreds of individual node-to-node connections is illustrated in Figure 1, since graphing each individual connection would lead to a blurred pattern of connecting lines. Instead a single line is used to stand in place of a fully interconnected pattern between levels. Learning is achieved by repetitive cycling through three steps. First, the system is presented with an input pattern that turns on some, but not all of the input units. In this case, the pattern is a set of sound features for the noun being used. Second, the activations of these units send activations through the hidden units and on to the output units. Third, the state of the output units is compared to the correct target and, if it does not match the target, the weights in the network are adjusted so that connections that suggested the correct answer are strengthened and connections that suggested the wrong answer are weakened.

MacWhinney et al. tested this system's ability to master the German article system by repeatedly presenting 102 common German nouns to the system. Frequency of presentation of each noun was proportional to the frequency with which the nouns are used in German {Baayen, 1993 #6855}. The job of the network was to choose which article to use with each noun in each particular context. After it did this, the correct answer was presented, and the simulation adjusted connection strengths so as to optimize its accuracy in the future. After training was finished, the network was able to choose the correct article for 98 percent of the nouns in the original set.

To test its generalization abilities, we presented the network with old nouns in new case roles. In these tests, the network chose the correct article on 92 percent of trials. This type of cross-paradigm generalization is clear evidence that the network went far beyond rote memorization during the training phase. In fact, the network quickly succeeded in learning the whole of the basic formal paradigm for the marking of German case, number, and gender on the noun. In addition, the simulation was able to generalize its internalized knowledge to solve the problem that had so perplexed Mark Twain -- guessing at the gender of entirely novel nouns. The 48 most frequent nouns in German that had not been included in the original input set were presented in a variety of sentence contexts. On this completely novel set, the simulation chose the correct article from the six possibilities on 61 percent of trials, versus 17 percent expected by chance. Thus, the system's learning mechanism, together with its representation of the noun's phonological and semantic properties and the context, produced a good guess about what article would accompany a given noun, even when the noun was entirely unfamiliar.

The network's learning paralleled children's learning in a number of ways. Like L1 German speaking children, the network tended to overuse the articles that accompany feminine nouns. The reason for this is that the feminine forms of the article have a high frequency, because they are used both for feminines and for plurals of all genders. The simulation also showed the same type of overgeneralization patterns that are often interpreted as reflecting rule use when they occur in children's language. For example, although the noun Kleid (dress) is neuter, the simulation used the overall sound form of the noun to conclude that it was masculine. Because of this, it invariably chose the article that would accompany the noun if it were masculine. Interestingly, the same article-noun combinations that are the most difficult for children proved to be the most difficult for the simulation to learn and to generalize to on the basis of previously learned examples.

How was the simulation able to produce such generalization and rule-like behavior without any specific rules? The basic mechanism involved adjusting connection strengths between input, hidden, and output units to reflect the frequency with which combinations of features of nouns were associated with each article. Although no single feature can predict which article would be used, various complex combinations of phonological, semantic, and contextual cues allow quite accurate prediction of which articles should be chosen. This ability to extract complex, interacting patterns of cues is a characteristic of the particular connectionist algorithm, known as back-propagation, that was used in the MacWhinney et al. simulations. What makes the connectionist account for problems of this type particularly appealing is the fact that an equally powerful set of production system rules for German article selection would be quite complex {Mugdan, 1977 #3030} and learning of this complex set of rules would be a challenge in itself.

8 Emergence in the Lexicon

One of the most active areas of current research in the child language is the study of early word learning. Philosophers like Quine {, 1960 #3407} have emphasized the extent to which word learning needs to be guided by ideas about what might constitute a possible word. For example, if the child were to allow for the possibility that word meanings might include disjunctive Boolean predicates {Hunt, 1962 #1979}, then it might be the case that the word “grue” would have the meaning “green before the year 2000 and blue thereafter”. Similarly, it might be the case that the name for any object would refer not to the object itself, but to its various undetached parts. When one thinks about the word learning task in this abstract way, it appears to be impossibly hard.

Markman {, 1989 #5719} and Golinkoff, Mervis, and Hirsh-Pasek {, 1994 #7517} have proposed that Quine's problem can be solved by imagining that the child's search for word meanings is guided by lexical principles. For example, children assume that words refer to whole objects, rather than parts of objects. Thus, a child would assume that the word "rabbit" refers to the whole rabbit and not just some parts of the rabbit. However, there is reason to believe that such principles are themselves emergent properties of the cognitive system. For example, Merriman and Stevenson {, 1997 #7518} have argued that the tendency to avoid learning two names for the same object emerges naturally from the competition {MacWhinney, 1989 #2725} between closely-related lexical items.

The idea that early word learning depends heavily on the spatio-temporal contiguity of a novel object and a new name can be traced back to Aristotle, Plato, and Augustine. Recently, Baldwin {, 1991 #5362;, 1989 #4705} has shown that children try to acquire names for the objects that adults are attending to. Similarly, Akhtar, Carpenter, and Tomasello {, 1996 #7523} and Tomasello and Akhtar {, 1995 #7524} have emphasized the crucial role of mutual gaze between mother and child in the support of early word learning. Moreover, Tomasello has argued that human mothers differ significantly from primate mothers in the ways that they encourage mutual attention during language. While not rejecting the role of social support in language learning, Samuelson and Smith {, 1998 #7541} have noted that one can also interpret the findings of Akhtar, Carpenter, and Tomasello in terms of low-level perceptual and attentional matches that help focus the child's attention to novel objects to match up with new words.

We can refer to the formation of a link between a particular referent and a new name as "initial mapping". This initial mapping is typically fast, sketchy, and tentative. Most lexical learning occurs after the formation of this initial mapping. As the child is exposed repeatedly to new instances of an old word, the semantic range of the referent slowly widens. Barrett {, 1995 #7528}, Huttenlocher {, 1974 #1992} and others have viewed this aspect of meaning growth as "decontextualization." Harris, Barrett, Jones, and Brookes {, 1988 #7529} have shown that the initial representations of words contain components that are linked to the first few contacts with the word in specific episodes or specific contexts. As long as the child sticks closely to attested instances of the category inside the confirmed core, she will tend to undergeneralize the word. For example the word "car" may be used to refer only to the family car. Anglin {, 1977 #100} and Dromi {, 1987 #7023} have argued that the frequency of such undergeneralizations is typically underestimated, because undergeneralizations never lead to errors. If one does a careful analysis of the range of uses of new words, it appears that undergeneralization is closer to the rule than the exception.

Gradually, the process of generalization leads to a freeing of the word from irrelevant aspects of the context. Over time, words develop a separation between a "confirmed core" {, 1984 #2703;, 1989 #2725} and a peripheral area of potential generalization. As the confirmed core of the meaning of a word widens and as irrelevant contextual features are pruned out, the word begins to take on a radial or prototype form {Lakoff, 1987 #2423;Rosch, 1975 #3519}. In the center of the category, we find the best instances that display the maximum category match. At the periphery of the category, we find instances whose category membership is unclear and which compete with neighboring categories {MacWhinney, 1989 #2725}.

According to the core-periphery model of lexical structure, overgeneralizations arise from the pressures that force the child to communicate about objects that are not inside any confirmed core. Frequently enough, children's overgeneralizations are corrected when the parent provides the correct name for the object {Brown, 1970 #550}. The fact that feedback is so consistently available for word learning increases our willingness to believe that the major determinants of word learning are social feedback, rather than innate constraints or even word learning biases.

One emergentist framework that allows us to model many of these forces is the self-organizing feature map (SOFM) architecture of Kohonen {, 1982 #5569} and Miikkulainen and Dyer {, 1990 #5381;, 1991 #4828}. These self-organizing networks treat word learning as occurring in maps of connected neurons in small areas of the cortex. Three local maps are involved in word learning: an auditory map, a concept map, and articulatory maps. Emergent self-organization on each of these three maps uses the same learning algorithm. Word learning involves the association of elements between these three maps. 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.

Evidence regarding the importance of syllables in early child language {Bijeljac, 1993 #7513;Jusczyk, 1995 #7515} suggests that the nodes on the auditory map may best be viewed as corresponding to full syllabic units, rather than separate consonant and vowel phonemes. The recent demonstration by Saffran et al. {, 1996 #7493} of memory for auditory patterns in four-month-old infants indicates that children are not only encoding individual syllables, but are also remembering sequences of syllables. In effect, prelinguistic children are capable of establishing complete representations of the auditory forms of words. Within the self-organizing framework, these capabilities can be represented in two alternative ways. One method uses a slot-and-frame featural notation from MacWhinney, Leinbach, Taraban, and

McDonald {, 1989 #2727}. An alternative approach views the encoding as a temporal pattern that repeatedly accesses a basic syllable map. A lexical learning model developed by Gupta and MacWhinney {, 1997 #6908} uses serial processes to control word learning. This model couples a serial order mechanism known as an "avalanche" {Grossberg, 1978 #6512} with a lexical feature map model. The avalanche controls the order of syllables within the word. Each new word is learned as a new avalanche.

The initial mapping process involves the association of auditory units to conceptual units. Initially, this learning links concepts to auditory images {Naigles, 1995 #7533;Reznick, 1990 #7534}. For example, the 14-month-old who has not yet produced the first word, may demonstrate an understanding of the word "dog" by turning to a picture of a dog, rather than a picture of a cat, when hearing the word "dog". It is difficult to measure the exact size of this comprehension vocabulary in the weeks preceding the first productive word, but it is probably at least 20 words in size.

In the self-organizing framework, the learning of a word is viewed as the emergence of an association between a pattern on the auditory map and a pattern on the concept map through Hebbian learning {Hebb, 1949 #1844;Kandel, 1992 #7022}. When the child hears a given auditory form and sees an object at the same time, the coactivation of the neurons that respond to the sound and the neurons that respond to the visual form produces an association across a third pattern of connections which maps auditory forms to conceptual forms. Initially, the pattern of these interconnections is unknown, because the relation between sounds and meanings is arbitrary {de Saussure, 1966 #3626}. This means that the vast majority of the many potential connections between the auditory and conceptual maps will never be used, making it a very sparse matrix {Kanerva, 1993 #6965}. In fact, it is unlikely that all units in the two maps are fully interconnected {Shrager, 1995 #7422}. In order to support the initial mapping, some researchers {Schmajuk, 1992 #5672} have suggested that the hippocampus may provide a means of maintaining the association until additional cortical connections have been established. As a result, a single exposure to a new word is enough to lead to one trial learning. However, if this initial association is not supported by later repeated exposure to the word in relevant social contexts, the child will no longer remember the word.

Parallel with the growth of the auditory map, the child is working on the development of an extensive system for conceptual coding. As we have noted, studies of concept development in the preverbal infant {Stiles-Davis, 1985 #6427;Sugarman, 1982 #4034;Piaget, 1954 #3295} indicate that the child comes to the language learning task already possessing a fairly well-structured coding of the basic objects in the immediate environment. Children treat objects such as dogs, plates, chairs, cars, baby food, water, balls, and shoes as fully structured separate categories {Mervis, 1984 #2923}. They also show good awareness of the nature of particular activities such as falling, bathing, eating, kissing, and sleeping.

Like auditory categories, these basic conceptual categories can be represented in self-organizing feature maps. Schyns {, 1991 #5492} applied a self-organizing feature map to the task of learning three competing categories with prototype structures. The individual exemplars of each category were derived from geometric patterns that were blurred by noise to create a prototype structure, although the actual prototypes were never displayed. The simulations showed that the network could acquire human-like use of the categories. When presented with a fourth new word that overlapped with one of the first three words, the system broke off some of the territory of the old referent to match up with the new name. This competitive behavior seems to reflect the process of competition between old words and new words discussed for children's word learning by Markman {, 1989 #5719}, Clark {, 1987 #809}, and MacWhinney {, 1989 #2725}.

Another simulation of meaning development by Li and MacWhinney {, 1996 #7113} used a standard backpropagation architecture to model the learning of reversive verbs that used the prefix "un-" as in "untie" or "dis-" as in "disavow". The model succeeded in capturing the basic developmental stages for reversives reported by Bowerman {, 1982 #460} and Clark, Carpenter, and Deutsch {, 1995 #6990}. In particular, the model was able to produce and later correct overgeneralization errors such as "*unbreak" or "*disbend". The network's eventual correct performance was based on its internalization of what Whorf {, 1941 #4391;, 1938 #4390} called the "cryptotype" for the reversive which involved a "covering, enclosing, and surface-attaching meaning" that is present in a word like "untangle", but absent in a form such as "*unbreak". Whorf viewed this category as a prime example of the ways in which language reflects and possibly shapes thought.

9 Emergence in Evolution

This section explores a very different type of emergentist account {MacWhinney, 2003 #9477}. Unlike the emergentist accounts reviewed in the previous two sections, this account cannot be directly implemented in mechanistic terms. Instead, it attempts to ground itself directly on known facts about brain structure, evolution, and language processing. In effect, this is a sketch of a class of possible emergentist accounts that must eventually be constructed in greater detail. This account is designed to link the emergence of language to specific evolutionary pressures that operated across the last 6 million years. These pressures are shown to have introduced a variety of modifications to cognitive structure that are in fact preconditions to language. Once these preconditions were in place, the final attainment of language was an emergent phenomenon.

MacWhinney {, 2003 #9477} analyzes the gradual evolution of language in terms of four major cognitive milestones. These four milestones are the buildling blocks of an embodied model of language processing developed in MacWhinney {, 1999 #7785}. That model views language as a method for taking a directly grounded perception and ungrounding it through imagery and perspective-switching.

9.1 Bipedalism

On the most grounded level, the model links language to cognition through the direct perception of affordances for action sequences. The second level links language to systems for spatial navigation and episodic encoding of temporal relations. These first two levels are associated first with the hominid assumption of a bipedal gait at about 4 MYA (million years ago). During the period between 4MYA and 2MYA, the model holds that our ancestors solidified the social role of language by linking vocal processes to cortical control. Beginning about 2MYA, homo erectus began to elaborate a mimetic system that provides the underpinnings for grammar. The introduction of a means for rapid control of phonation at about 200,000 years ago then led to a linkage of these earlier cognitive systems to a full system for using language to control social interactions.

This account emphasizes the role of specific neuronal adaptations at each of these evolutionary junctures. The move to bipedalism opened up major cognitive challenges in terms of the control of the hands. Apes already have good control of reaching and basic object manipulation {Ingmanson, 1996 #9434}. However, with both hands now always free for motion, humans were able to explore still further uses of their hands. Rizzolatti {, 1996 #7776} has shown that monkeys (and presumably also primates) have “mirror” neurons in the supplementary eye fields of premotor cortex that allow them to directly map their own body image onto that of a conspecific. The basic neural mechanism for assuming the postural perspective of another would allow an early hominid to directly track and imitate the motions of other hominids. It allows them to follow actions such as prying open shells, hitting things with clubs, and digging for roots.

The construction of a mental image for controlling motor plans depends on the dorsal visual pathway that processes actions upon objects {Goodale, 1993 #7687}. As hominids increased their ability to control hand motions and grasping actions, they could use this system to link specific actions to the affordances of different objects, as they are used for different purposes. The move to a terrestrial environment was quite gradual {Corballis, 1999 #9464}. This meant that hominids needed to provide neural control for both tree-climbing activities and the use of the arms when walking bipedally on the ground. The pressures in the arboreal environment that favor some limited form of brain lateralization were then carried over to the terrestrial environment {McManus, 1999 #9465}. This ability to shift quickly between alternative environments required neural support for competing postural and affordance systems. This postural flexibility may also have allowed some early hominids to adapt partially to an aquatic environment {Morgan, 1997 #9459}.

Bipedalism also put some pressure on another set of neural mechanisms. Because hominids ceased relying on trees for refuge, and because they were now ranging over a wider territory, they needed to develop improved means of representing spaces and distances. Holloway {, 1995 #9426} has presented evidence from endocasts indicating that there was, in fact, a major reorganization of parietal cortex after about 4 MYA. This reorganization involved the reduction of primary visual striate cortex and the enlargement of extrastriate parietal cortex, angular gyrus, and supramarginal gyrus.

9.2 Cortical control of the vocal system

The second major reorganization of cognitive functioning introduced cortical control over the vocal-auditory channel. As Holloway {, 1995 #9426} has stressed, this change does not require a major increase in brain size. However, it does require a rather major rewiring of the relation between frontal cortex and the limbic system. In macaques {Jürgens, 1990 #9510}, control of the vocal system relies on the periaqueductal gray matter of the lower midbrain. Additional midbrain regions can stimulate the periaqueductal gray, but the cortex does not control or initiate primate vocalizations. In man, on the other hand, electrical stimulation of both the supplemental motor area and the anterior cingulate of the frontal cortex can reliably produce vocalization. Tucker {Tucker, 2003 #9485} shows that the basic adaptation here involved the absorption of the primate external striatum by the neocortex {Nauta, 1970 #9446}.

The linkage of vocalizations to cortical control not only allowed our ancestors to distinguish themselves from other hominids, it also allowed them to build up a system of face-to-face social interactions. MacNeilage {, 1998 #8621} has argued that the primate gesture of lip smacking is the source of the core CV syllabic structure of human language. The CV syllable has the same motoric structure as lip smacking and its is produced in an area of inferior frontal cortex close to that used for lip smacking and other vocal gestures. Primates use lip smacks as one form of social interaction during face-to-face encounters. However, even bonobos, the most social of all primates, do not maintain face-to-face conversations for the long periods that we find in human interactions. By linking its members into tight affiliative relations through face-to-face interaction, our ancestors achieved a form of social organization that allowed them to maintain large social groups for defense against other hominid groups. Other primates have also responded to these pressures by developing a variety of social support mechanisms {de Waal, 1996 #9462}. Other primates have also developed systems for attending to face-to-face interactions and pointing behavior {Gomez, 1996 #9463}. To maximize the effectiveness of face-to-face interactions, hominids then brought the production of facial gestures under cortical control. As in the case of the control of tool use through motor imagery, humans differ from monkeys in the extent to which the cortex can produce gestures upon demand {Myers, 1976 #9440}.

In considering the role of face-to-face vocalization in hominid groups, we must not forget the possible divisive role played by aggressive males {Anders, 1994 #9457;Goodall, 1979 #9456}. Hominid groups relied on aggressive males for their skills as hunters and their ability to defend the group against attack. However, groups also needed to provide ways to avoid the direction of male aggression toward other members of the group, particularly other males. We know that primates had already developed various methods for handling these conflicts, including exile for problematic males, the formation of master-apprentice relations, and development of male social groups. Within this already established social framework, males could also benefit from ongoing reaffirmation of their social status through face-to-face chat. By socializing young males into this productive use of language for social cohesion, mothers could also contribute to the stability of the group. Breakdowns in these processes could threaten the survival of the group and even the species.

9.3 Mimesis

The brain size of homo erectus tripled in size during the period between 2 MYA and 100,000 years ago. This growth reflects the growing importance of protolanguage in homo erectus or homo ergaster. In order to maintain this mimetic system, these neuronal adaptations were required:

1. The production system must link up stored visual representations to the output processes of chant, gesture, and dance. This linkage of vision to gestural and vocal output requires not only the expansion of both central and peripheral {MacLarnon, 1999 #9425} output control areas , but the expansion of their connections to basic visual areas {Givón, 1998 #9410}. More generally, control of this system requires the construction of a cognitive simulation of the human body {MacWhinney, 1999 #7785}.
2. Mothers must be able to socialize their children into an understanding of the core mimetic sequences of their own social group.
3. The episodic memory system must store mimetic sequences and their components.
4. As mimetic sequences become elaborated, the brain will need to provide methods for storing whole perspectives, such as that of the hunter, to allow for a switching of perspective, as well as traditional reenactment of these shifts.

Unlike the evolutionary pressures of earlier periods, the cognitive pressures imposed by mimesis cannot be solved simply by linking up older areas or by reusing earlier connections. Instead, the brain must add new computational space to store the multitude of new visual and auditory images {Li, 2003 #9486}. In addition, the brain needs to expand the role of the frontal areas for storing and switching between perspectives. Because this system grew up in a haphazard way from earlier pieces of lip smacking, pointing, gesture, and rhythm, the brain cannot simply extract a core set of elements from mimetic communications, thereby reducing requirements for storage space. Instead, many patterns and forms must be learned and stored as whole unanalyzed sequences. This Gestalt-like shape of early mimetic patterns corresponds well with the Gestalt-like cognitions that we develop through our interactions with objects. For example, when we chop wood, there is a complete interpenetration of muscle actions, visual experiences, hand positions, and sounds. We can think of this as a single merged form such as I-hands-back-lift-axe-drop-split-chips-wood-cut. Mimetic forms have this same unanalyzed quality. Because they are highly grounded on our direct perceptions and actions, they communicate in a basic way. However, they provide little support for cognitive organization.

The growth of the brain in response to these pressures was so rapid that it is typically assumed that it involves a single genetic mechanism. One such mechanism might well be the role of a few regulatory genes {Allman, 1999 #9525} in controlling the overall size of the cortex. Changes in the function of these genes can lead to the observed across-the-board increase in size for the cortex and cerebellum that we see in *homo erectus*. However, the expansion of the cortex placed additional adaptive pressures on *homo erectus*. One was the need to increase caloric intake to support the metabolic needs of a larger brain. This pressure could be met through changes in diet and modifications to the digestive system. A more fundamental pressure was the fact that increases in the size of the infant brain produce problems for the process. The width of the hips had narrowed in both men and women as a response to bipedalism. As long as the skull was not much larger than that found in the primates, this did not cause major problems. However, the expansion of the skull in *homo erectus* ran directly into this evolutionary barrier. To deal with this, the infant is born at a time when it is still fairly immature and the skull is relatively pliable. The increasingly organized shape of the society guarantees the survival of the child. In addition, women have had to sacrifice their ability to run quickly so that the hips could widen, permitting births with larger infant heads. The slowing of infant development not only helps in the birth process, but also helps the child maintain cortical plasticity {Elman, 1996 #7691;Julész, 1995 #7413} even into adolescence, thereby further enhancing the ability of the group to construct accepted mimetic patterns.

9.4 Systematization

Some of the adaptations required for smooth vocal production are quite peripheral. {Lieberman, 1973 #2565}, involving changes to the vocal tract, the structure of the larynx, muscle innervation, tongue support, and facial musculature. Some of changes were underway before the Pleistocene; others have been more recent. To control this additional external hardware, the brain has needed to fine-tune its mechanisms for motor control. This fine-tuning does not require the type of brain expansion that occurred in *homo erectus*. Instead, it involves the linking of inferior frontal areas for motor control to temporal areas {Gabrieli, 1997 #7658} for sequence storage. These linkages {Damasio, 1989 #7156} involve pathways that lie under the central sulcus. They constitute a functional neural circuit that implements a phonological loop for learning new words {Gupta, 1997 #6908}. The auditory shapes of words are stored in topological maps {Miikkulainen, 1990 #5381} in superior temporal auditory cortex and can be associated to visual images in inferior temporal areas. This linkage of the vocal-auditory channel to the visual channel further develops binding the entrainment of the vocal-auditory channel by the visual channel {Givón, 1998 #9410}.

Once homo sapiens had achieved an ability to produce, store, and learn a large vocabulary of phonologically organized forms {Wode, 1994 #9526}, the remaining steps in the evolution of language were comparatively easy. Humans had already achieved a mimetic system for perspective taking and perspective-switching. This system allowed listeners to mentally reenact the motions, rhythms, and chants of the speaker as they depicted movement between places and actions on objects. Once words became available, speakers and listeners could parse these single-package gestalt-like communications into their components. With words to name specific objects and participants, it was possible to separate out nouns from verbs. This adaptation to grammar required no particular new cognitive skill for nouns. However, for predicates such as verbs, it was important to store linkages between the overall configuration of the action and the specific uses with participants. In other words, children had to learn how to manage language in terms of item-based syntactic constructions {MacWhinney, 1982 #2699;MacWhinney, 1975 #2683}, including “verb islands” {Tomasello, 2000 #9266}. Neuronal processes for this level control involve little in the way of new evolution. However, they place storage demands on the pathways between the temporal lexical areas and the frontal planning and sequencing areas.

As speakers build up longer and longer strings of propositions, they rely increasingly on frontal areas, such as dorsolateral prefrontal cortex (DLPFC) for the storage of one perspective that allows shifting to a secondary perspective. Shifts of this type are central in the processing of anaphors and gaps in argument structure. As MacWhinney {, 2000 #9255} has shown, these various syntactic processes are grounded not on the construction of abstract syntactic trees, but on the direct processing of embodied perspectives of the type that were also important during the period of mimetic communication.

Given the minimal nature of the additional adaptations needed to support human language, why was the human species suddenly so successful after about 60,000 years ago after having lived through near extinction? The reason for this great success is that, with the onset of good phonological systematization, humans were able make a fuller use of the massive expansion in brain size that had occurred earlier. They did this by constructing a system that uses the entire brain to represent experience. It uses the basic sensorimotor systems of posterior cortex to encode objects in terms of direct perceptions and properties. It uses the navigation system that developed in the hippocampus and the temporal lobe to organize deictic terms, prepositions, and locative adverbs. It relies on the system that attributes intentional to conspecifics to construct causal actions by both animate and inanimate actors. It uses the temporal and inferior frontal areas to encode the form and meaning of vocal-auditory patterns. It then relies on a wide variety of frontal structures to store and shift perspectives in terms of direct perception, spatial systems, and causal actions. Finally, it uses this system of perspective shifting in the frontal lobes to construct the complexities of social structure. In effect, language production and comprehension end up relying on the entire brain. In this way, the phonological systematization that occurred between 200,000 and 50,000 years ago eventually succeeded in utilizing the full potential of the earlier expansion of the brain.

Language relies on the entire brain to achieve its complete cognitive simulation of experience in terms of objects, space, action, and social relations. Because it integrates these separate modules so thoroughly, it allows us to fully escape the modularity that is present in primates {Russon, 1996 #9451} and young children {Hermer-Vazquez, 2001 #9388}. Without language, it may be possible to focus directly on the position of an object without regard to earlier orientations or the orientations of others. Without language, we can focus on an action without breaking it apart into its component participants. Without language, we are more directly grounded in the individual aspects of mental life. Language forces us to integrate the whole of mental life into a single, more fully conscious, but relatively less grounded whole.

10 Conclusion

The core lesson of the last century is that both empiricism and nativism are wrong. Empiricism is wrong because it attempts to construct the mind out of nothing but domain-general 'buzzsaws.' Nativism is wrong because it makes untestable assumptions about genetics and unreasonable assumptions about the hard-coding of complex formal rules in neural tissue. The battles against disembodied behaviorism were fought and won in the 1950s. The battle against complex strictly-ordered rule systems was fought and won in the 1980s. We have made great progress and these issues are no longer on the table. Emergentism provides a stable roadmap for further scientific progress that replaces the battles of the last century with the hard work of formulating concrete mechanisms for structural emergence.

References: