The Origins of Arbitrariness in Language

Michael Gasser (gasser@indiana.edu)
Computer Science Department; Lindley Hall 215
Indiana University; Bloomington, IN 47405 USA

Abstract

Human language exhibits mainly arbitrary relationships between the forms and meanings of words. Why would this be so? In this paper I argue that arbitrariness becomes necessary as the number of words increases. I also discuss the effectiveness of competitive learning for acquiring lexicons that are arbitrary in this sense. Finally, I consider some implications of this perspective for arbitrariness and iconicity in language acquisition.

A Language Design Task

Imagine you are inventing a language. It should associate signals ("forms") that can be produced and perceived by the users of the language with perceptual or motor categories ("meanings"). Assume that both forms and meanings are patterns of values across sets of dimensions and that you have been given the form and meaning dimensions. Assume further that the specific design task includes a set of meaning categories that need to get reliably conveyed. That is, given a particular pattern across the meaning dimensions, if it belongs to one of the given set of categories, a user who knows the language should be able to assign a form to it, that is, an appropriate pattern across the set of form dimensions. Similarly, given a pattern across the form dimensions, if it belongs to one of the set of form categories that you have built into your language, a user who knows the language should be able to assign a meaning to it. Furthermore, the form assigned to an input meaning should be the "right" form; that is, the form that gets output should pass the comprehension test in the reverse direction. Providing this form to a user who knows the language should result in an output meaning that is at least closer to the original meaning than to any of the other meaning categories. In the same fashion, the meaning assigned to an input form should pass the production test in the reverse direction.[1]

Your language is not hard-wired into a user; it must be learned through a series of presentations. A presentation consists of a pairing of a form and a meaning selected randomly from the set of possible form-meaning pairs that are built into the language, with a small amount of noise added to both the form and the meaning. Two constraints that you need to consider in your design are ease of learning and ease of storage. Each user has finite resources for learning and storage, and there is an advantage to languages that are learned with fewer presentations.

The main issue of concern in this paper is how the solution to a language design task of this type is constrained by the number of distinct meanings that are to be conveyed by the language. I will argue that there are advantages to languages with systematic relationships between forms and meanings and advantages to languages without such systematicity. I will then discuss how competitive learning fares at learning both types of languages. Finally I will discuss the implications for acquisition and evolution of human language.

Iconicity and Arbitrariness

How Iconicity Can Help

Learning the association between forms and meanings can be facilitated if there is a systematic relationship between the patterns. A simple example of such a relationship is a correlation between the values on a form dimension and a meaning dimension. There are two possibilities for where such a correlation might come from. One is for it to be based on a natural relationship between the two dimensions, for example, if they are the same dimension at a more abstract level. Such relationships are familiar in human language from onomatopoeia, in which form imitates meaning on one or more acoustic/auditory dimensions, for example, pitch. Examples of this type are more common in sign languages, where a movement of the hand in signing space may represent a physical movement of some object in meaning space.

A further possibility is for the relationship between the correlating dimensions to be completely arbitrary, or at least opaque to the users. In some sign languages, for example, American Sign Language and Japanese Sign Language, movement towards or away from the head represents the gain or loss of knowledge: learning, remembering, forgetting. But the motivation for the association between the form and meaning dimensions in this case would require that the user know that knowledge is in some sense in the head. Thus the relationship between the form and meaning dimensions in this case could be viewed as arbitrary by a particular learner, though the learner might still notice the systematicity of the rela-

[1]Note that in this sense, these simple languages deviate from human languages, which permit multiple forms for the same meaning and multiple meanings for the same form. But the constraint has to roughly hold for communication to get off the ground, and young children learning language seem to behave as though it does (Markman, 1989).

tionship, that is, that within this set of signs the head represents the location of knowledge.

These kinds of systematic relationships between form and meaning are referred to as **iconicity**. I'll return to the topic of iconicity and **arbitrariness**, the absence of iconicity, in human language later in the paper.

In an iconic language, there is less to learn than in a purely arbitrary language, so learning should be faster and require less storage. This is easily seen by imagining a language with five meanings to be conveyed and a single dimension each for form and meaning. An arbitrary language would require storing separately each of the five form-meaning pairs of values on this dimension, but a completely iconic language with a perfect correlation between the dimensions would only require a single value, a correlation of 1.0. This is illustrated in Figure 1.

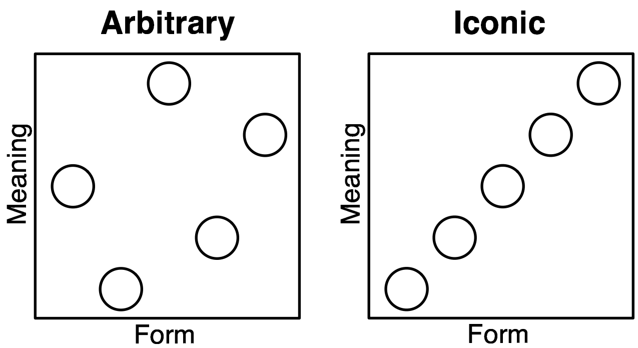

Figure 1: Arbitrariness and iconicity. Two simple languages, each with one form and one meaning dimension and five meanings to be conveyed. Noisy form-meaning pairs are indicated by circles in form-meaning space. In an arbitrary language, there is no correlation between form and meaning. In a perfectly iconic language, form and meaning correlate.

Iconicity can play a further role in the comprehension of the language. If an unknown or poorly learned form is presented in the presence of constraints on the possible meanings for the form, for example, if several candidate meanings are present, then iconicity can add further constraints. For example, if a user of a language knows that loudness in forms that refer to emotions tends to correlate with the strength of the emotion referred to, then for a particularly loud novel form, the user can eliminate candidate emotions that are mild.

How Iconicity Can Interfere

However, this advantage of iconicity should decline as the number of meanings to be associated with forms increases. Increasing the number of form-meaning pairs causes the average distance between these pairs in form-meaning space to decrease. Because of the noise that is part of form and meaning patterns, each form-meaning association occupies a region of the space. In other words, as the number of form-meaning pairs increases, the likelihood that the form regions for two different pairs share the same meaning (homophony) or that the meaning regions for two different pairs share the same form (ambiguity) increases. Obviously both sorts of overlap can interfere with communication; a noisy form pattern might get assigned to more than one meaning category, for example. They also interfere with learning; it will be more difficult to make the proper associations if forms or meanings are sometimes ambiguous.

Now consider how iconicity affects the likelihood of these sorts of overlap. Because iconicity constrains the possible form-meaning associations, it results in a narrowing of the space. This is illustrated in Figure 2. If we imagine the fixed set of meanings that are to be conveyed in the language as non-overlapping channels in the form-meaning space, then the possible forms for each can be viewed as circles (or hyperspheres in spaces of more dimensions) that can be slid back and forth in the channels, resulting in different languages. If we arrange two of these circles so that a portion of one is above a portion of another, we have the sort of overlap that represents ambiguity. There are obviously more ways to arrange the circles and avoid ambiguity in an arbitrary language like the one on the left than there are in a highly iconic language like the one on the right.

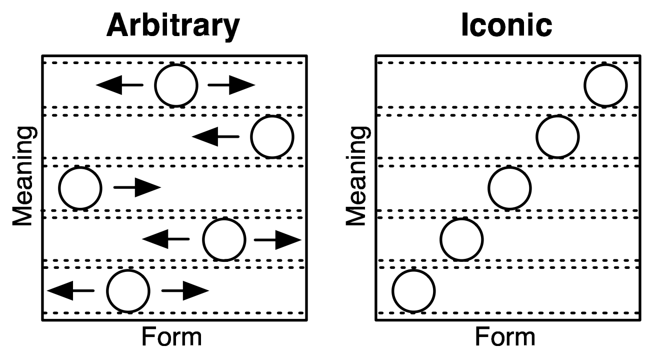

Figure 2: Arbitrariness, iconicity, and vocabulary size. For relatively large vocabularies, iconicity can interfere with communication because of the greater likelihood of overlap between form-meaning pairs. For a given vocabulary size, there are more ways to avoid ambiguity (and homophony) in an arbitrary than an iconic language.

A Simulation

For a learning algorithm that responds to regularities in the association between form and meaning, then, we should observe an interaction between vocabulary size and systematicity in the association (arbitrariness vs. iconicity), as measured by learning error.

To test this idea, I trained several feedforward connectionist networks to learn the associations from a set of meanings to a set of forms. The languages differed on two dimensions, vocabulary size and systematicity in the association. Both forms and meanings were represented by values along three dimensions, with ten possible values for each. Each dimension was represented by ten units, and each input and target value activated a gaussian pattern across the units so that there was the possibility of some generalization from a value to values close to it.

The "small" languages contained 15 form-meaning pairs, while the "large" languages contained 100 form-meaning pairs. For "iconic" languages, each form-meaning pair coincided on two of the three dimensions, which were randomly selected for each pair. For example, a possible iconic form-meaning pair was: form {3, 2, 8}, meaning {3, 5, 8}. Note that for the iconic languages, there is thus a significant correlation across all three pairs of dimensions. For "arbitrary" languages, the values for each form-meaning pair were selected completely randomly. For each form-meaning pair, the network saw five separate presentations, one with the canonical pair, and four with noisy variations on this pair. For each of these variations, each dimension value was changed by 1 with a probability of 0.2.

Since these were feed-forward networks, they only learned the associations in one direction. Each network contained 30 meaning input units, 30 form output units, and 64 hidden units and was trained using back-propagation. Figure 3 shows the mean square error as training progressed. As can be seen, iconic languages have an early advantage because of the correlations that back-propagation can easily discover. For the small languages, this advantage holds throughout training. For the large languages, however, the network learning the arbitrary language eventually overtakes the one learning the iconic language, apparently because of the proximity of some of the form-meaning pairs to one another and the resulting confusion in the presence of noise.

Note that the potentially adverse effects of iconicity on learning depend crucially on the number of dimensions that are used to represent forms and meanings because the size of the form-meaning space increases with the number of dimensions. For a large enough number of dimensions, iconicity should be superior to arbitrariness, even for a relatively large vocabulary. In fact, if we increase the number of dimensions in the simulation from three to four, the long-term advantage of the arbitrary over the iconic language disappears.

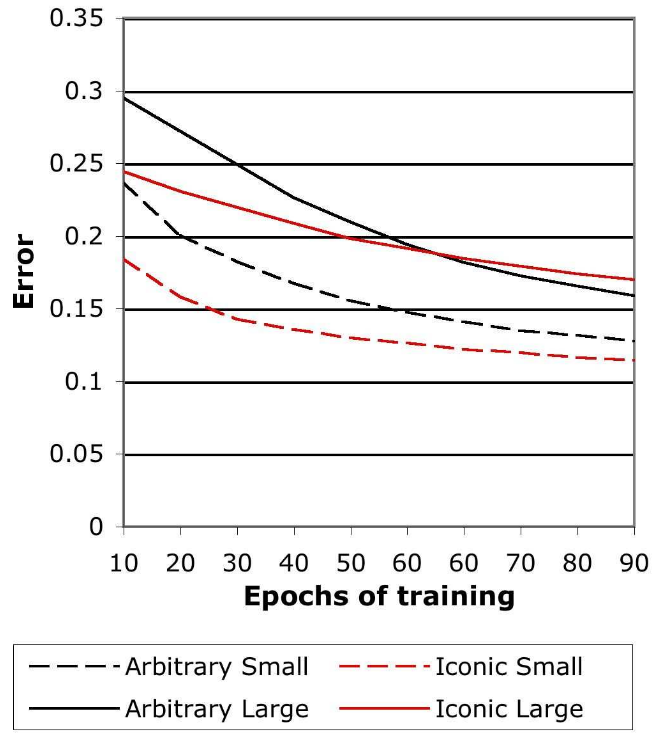

Figure 3: Learning of iconic and arbitrary languages by a feed-forward network. Root mean square error during training is shown for iconic and arbitrary languages consisting of 15 (small) and 100 (large) form-meaning pairs.

Arbitrariness and Competitive Learning

Learning Arbitrary Categories

Let us assume that the communicative demands of the users of the language require forms for a very large number of meanings and that the number of form and meaning dimensions available for representing forms and meanings is small enough that a mostly arbitrary language has a clear advantage over a mostly iconic one.

Now suppose we have some control over the kind of learner that is confronted with this large and mostly arbitrary language. What sort of learning mechanism would be best suited for this task? What matters most is that the different form-meaning pairs be kept distinct from one another. That is, each of these is in effect a separate category. (Since we are now dealing with categories of form-meaning association, it is time to start calling them "words.") Since in an arbitrary language there is little or no regularity to be found between the categories, an algorithm that focuses on within-category regularity, while it ignores between-category regularity, makes sense. Of course, the categories are not specified to the learner in advance; the learner neither knows how many form categories there are nor how many meaning categories there are. Thus the algorithm must be unsupervised.

Competitive learning (e.g., Grossberg, 1987) is such an algorithm (or family of algorithms). It seeks to cluster input patterns on the basis of similarity, and it is oblivious to any regularities that exist between the categories that it finds. It would seem to be well-suited to the task of learning words. But how does it respond to iconicity and arbitrariness?

A competitive learning network has an input layer and an output layer consisting of potential category units. The output layer either has a fixed number of units, representing an upper bound on the number of categories that can be learned, or, in a constructive competitive learning algorithm, the output layer adds new category units in response to error. In the simple version of competitive learning used here, for each input pattern the category unit whose weights best match the input pattern is treated as the "winner" for that pattern. It updates its input weights in the direction of the input pattern. The "losing" units also update their weights in the direction of the input, but with a much smaller step size.

A competitive learning network for the form-meaning learning task has both form and meaning as inputs feeding into an output layer of category units. During training, an input pattern consisting of a form-meaning pair activates a winning unit, and the weights are updated. Ideally, a single category unit gets assigned to each form-meaning category; that is, each unit ends up representing a word. A single training presentation is illustrated in Figure 4A.

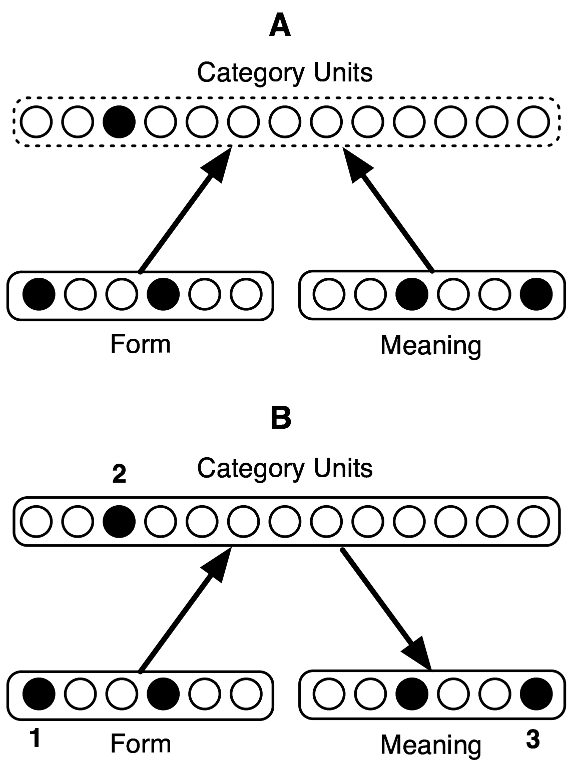

Figure 4: Competitive learning of form-meaning pairs. A. Training. An input pattern, consisting of both form and meaning patterns, is presented to the network, which selects a "winning" category unit, and updates its weights and, to a lesser extent, the weights of other units. In the constructive version of the algorithm used in the simulation, the category layer grows during training (indicated by the dashed border); it adds a new unit whenever error for an input pattern is above a threshold. B. Comprehension. An input pattern, consisting of a form pattern only, is fed to the network (1) and the winning category unit is activated (2). The active category unit activates a pattern on the meaning units (3).

Following training, the network can perform production or comprehension using the trained weights. For comprehension, a form pattern alone is input to the network, and a winning category unit is selected on this basis. This unit then activates the meaning units using the weights learned in the other direction. Production works in the opposite fashion, with meaning as input and form as output. Figure 4B shows how comprehension is implemented.

A Simulation

To test whether competitive learning could elucidate both the advantages and disadvantages of iconicity, I trained a competitive learning network of the type described above on both a completely arbitrary language and a maximally iconic language, in which all form dimensions correlated with meaning dimensions. There were four meaning and four form dimensions and 100 form-meaning pairs in the language. In addition to the form and meaning input layers, the network had a growable layer of category units. At each input presentation, a new category unit was added with a probability based on the error for the input pattern (the distance of the winning category unit from the input). Separate identical networks were trained for 50 epochs on the two kinds of languages. Figure 5 shows the results for several kinds of tests following training.

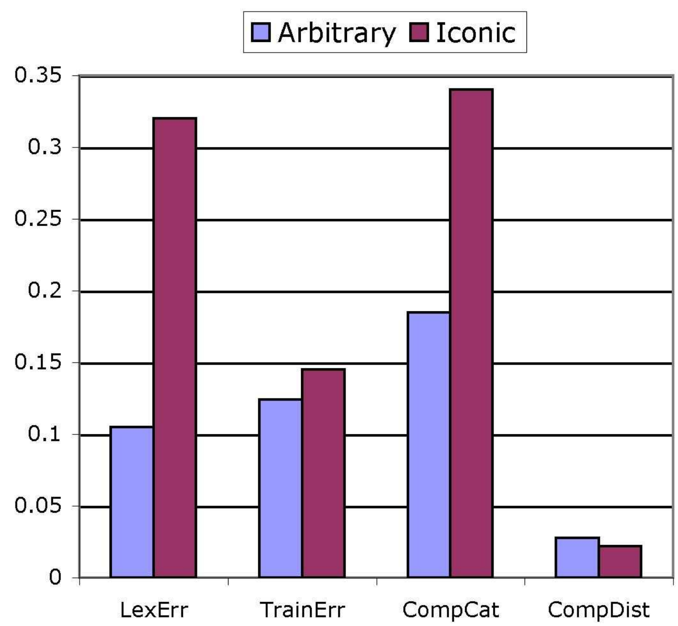

Figure 5: Competitive learning of arbitrary and iconic languages. Results are shown for the proportion of words that are not assigned distinct category units ("LexErr"); the final error on training patterns, that is, the average distance of input patterns from winning category units ("TrainErr"); the proportion of words in comprehension tests for which the meaning output was closer to a meaning category other than the intended one ("CompCat"); and the average distance of the meaning output in comprehension tests from the intended meaning ("CompDist").

The first two columns show tests directly related to the degree to which the networks mastered the languages. The first column shows what proportion of the 100 words became associated with distinct category units during training. Any category unit that ends up representing more than one word will obviously interfere with comprehension or production. For the iconic language there are more units doing double duty because of the greater similarity between the words. The second column shows another measure of learning, the average distance between an input pattern and the category unit that wins when it is presented following training. The smaller this number, the more successful the network has been in handling all of the words. Again the network trained on the arbitrary language out-performs that trained on the iconic language.

The third and fourth columns represents tests of comprehension of forms, one for each of the words in the training set. There are two ways to test comprehension. One determines whether the meaning that is output is closer to the intended meaning (the one actually associated with the form in the language) than to any other. The result for this test appears in the third column. Again the arbitrary language has an advantage. A second way to test comprehension measures the distance between the meaning that is output and the intended meaning. The result for this test appears in the fourth column. Here the iconic language has a small advantage, one that holds over a range of parameter settings. This can be explained by considering what happens when a noisy or poorly learned form is presented to a network that has learned the iconic language. Even if the category unit that wins for this input is not the appropriate one, that is, the one that would yield the intended meaning, the meaning that is output will not be far off. Somewhat surprisingly, then, even though the iconic language is less well learned, it is more easily comprehended in this sense.

Human Language

What does all of this have to do with human language? Since at least the work of de Saussure (1983), it has been recognized that the association between form and meaning in human language is largely arbitrary. However, in Saussure's work and in other influential work by scholars such as Peirce (1998), iconicity and arbitrariness seem never to have been spelled out clearly enough to admit to any sort of rigorous test. They have always boiled down to "motivation" or "resemblance" or their absence.

The discussion above provides both a formalization of iconicity and arbitrariness and an account of why human language might have a strong tendency to be arbitrary. For whatever reason, we need to distinguish tens of thousands of categories of objects, attributes, states, and events, and the associations between these categories and the forms that convey them in a language need to be stored in a brain and to be learned through presentations that do not make explicit what the categories are. Under these circumstances, the arbitrariness of the form-meaning association helps keep words separate during learning.

Another implication of the discussion above is that word learning and word access in humans is a competitive process, that words are categories. This isn't a novel idea at all. In fact models of word recognition (e.g., Norris et al., 2000) and word access in language production (e.g., Levelt et al., 1999) that are not competitive are the exception. And the fact that competitive learning results in localized representations of words is compatible with the idea that words are the origin of symbolic behavior (Vygotsky, 1978).

But this brings up more questions. First, what about iconicity in human language? It is well-known that, far from being non-existent, iconicity actually thrives in some corners of language (Hinton, Nichols, & Ohala, 1994). It is a property of so-called expressive words, which make up an entire grammatical category in a wide range of languages, including Japanese, Korean, and many languages spoken in Africa, South Asia, Southeast Asia, and the Americas. It is also much more common in sign languages (Taub, 2001) than in spoken languages.

Given what I have claimed, we would expect iconicity in circumstances where the number of words is unusually small or in circumstances where the space of possible distinguishable forms is unusually large. The number of words is small early in first language acquisition, and there is some evidence that in at least one language with a large category of iconic words, Japanese, these words are relatively common in speech to children and they are easier for children to map onto meanings than arbitrary forms are (Yoshida, 2003). That is, they seem to play the role in comprehension that is suggested by the discussion above. Another situation in which a vocabulary is very small is experiments in which which adults have to communicate with one another without speaking. Not surprisingly, subjects in such experiments create highly iconic gestures to represent categories of objects and relations (Oda & Gasser, 2003).

However, expressives survive into the adult language for speakers of languages like Japanese, Tamil, and Zulu. One possible explanation is that these categories are more or less self-contained, existing in a sense in their own space. They tend to be characterized by particular formal properties such as reduplication, and they tend to convey particular categories of meanings such as movements, sounds, and textures. Perhaps expressives fail to interfere with other words because learners place them in a category all by themselves.

But what of sign languages? Although there is no evidence yet that the iconicity of sign languages helps young children pick up the meanings of words, there is lots of anecdotal evidence that adults learn sign languages relatively rapidly, presumably because of the iconicity. But how can we account for the pervasiveness of iconicity in the vocabularies (not to mention the grammars) of these languages? Although there is an apparent tendency towards somewhat less iconicity as these languages change, there is no evidence that the iconicity is disappearing (Taub, 2001). Of course it is possible that sign languages are more iconic than spoken languages because there are more ways to be iconic in the spatial than in the acoustic domain. But that does not explain how all of the iconicity can be tolerated, how the words keep from overlapping in the sense I have discussed. One possibility suggested by the account I've sketched is that the space itself is larger, that the number of dimensions along which signs vary or the number of distinguishable values along these dimensions is greater than it is for spoken word forms. This seems worth investigating.

Finally, how would competitive learning deal with a language, or a subset of a language, that exhibited some iconicity, along with the more normal arbitrariness? The competitive network discussed above is doomed to being thrown off by the iconicity. Although it might, in the short run, perform better on a comprehension task, as happened in the simulation above, in the long run, it needs to be able to keep words separate from one another. However, there is nothing about competitive learning that restricts it to a single layer of category units. A more flexible network in fact is one that allows for different degrees of granularity in how the clustering of inputs takes place. This is achieved with layers with different numbers of category units or, for constructive networks, with different thresholds for the creation of new category units. The competition among units to classify inputs is only within, not between the layers. Such a network is shown in Figure 6. A network like this was trained on a set of 100 words, again with four form and four meaning dimensions, in which either the first form and first meaning dimension correlated or the second form and second meaning dimension correlated. The other two dimensions of form and meaning were uncorrelated. The larger category layer learned the set of words as before (note that the behavior of this layer is completely unrelated to the behavior of the other), while the smaller layer divided the patterns into two clusters. A comprehension task in a network like this relies on two winning category units, rather than one. That is, it can combine the correlational information embodied in the weights to the smaller layer with the arbitrary associations embodied in the weights to the larger layer.

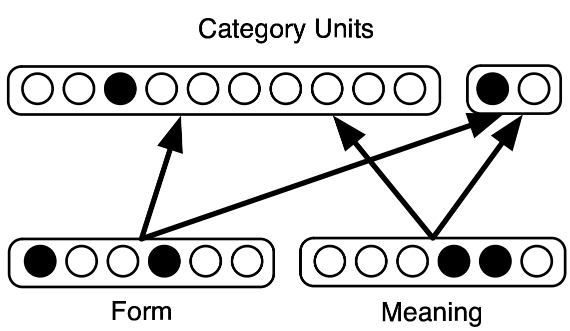

Figure 6: Competitive learning with multiple layers of category units. The number of category units (or the threshold for the creation of new units) in a layer governs the number of categories that it discovers.

Conclusions

Since iconicity seems to make so much sense and since humans are so good at imitation, it might seem surprising that human languages exhibit such overwhelming arbitrariness in the form-meaning relationships that define words. I have tried to show in this paper how the sheer number of concepts we feel the need to talk about inhibits us from making use of this strategy. It's crucial that words be kept separate, and it's easier to do this if there's little or no sense to how forms relate to meanings. This arbitrariness in turn favors algorithms that categorize form-meaning pairings, in short, algorithms that learn words. On this view, words are the local representations that result from the competitive learning of mainly arbitrary form-meaning associations.

But if this is so, how did it or how does it get that way? Did the advantage of being a competitive learner of form-meaning pairings cause our ancestors to evolve this approach to language? Or is this a mechanism that develops in children as they are exposed to a system that mostly fails to be iconic? Investigating the first possibility using evolutionary algorithms and investigating the second through the modeling of early word learning in children are future directions for this project.

References

de Saussure, F. (1983). *Course in General Linguistics*. G. Duckworth, London.

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

Hinton, L., Nichols, J., & Ohala, J. (Eds.). (1994). *Sound Symbolism*. Cambridge University Press, Cambridge.

Levelt, W. J. M., Roelefs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. MIT Press, Cambridge, MA.

Norris, D., McQueen, J. J., & Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.

Oda, H. & Gasser, M. (2003). Gesture Language Game — simulating the emergence of linguistic signs.. Paper presented at the 8th International Cognitive Linguistics Conference, Logroño, Spain.

Peirce, C. S. (1998). *The Essential Peirce: Selected Philosophical Writings: Volume 2*. Indiana University Press, Bloomington, IN.

Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press, Cambridge.

Vygotsky, L. S. (1978). *Thought and Language*. MIT Press, Cambridge, MA.

Yoshida, H. (2003). *Iconicity in Language Learning: the Role of Mimetics in Word Learning Tasks*. Ph.D. thesis, Indiana University, Bloomington, IN.