An Architecture for the Learning of Perceptually Grounded Word Meanings

Michael Gasser and Eliana Colunga
Computer Science Department, Cognitive Science Program
Indiana University
Bloomington, Indiana 47405
gasser@cs.indiana.edu, ecolunga@cs.indiana.edu

In this statement, we discuss two kinds of properties that a grounded model of the learning of word meaning should have, those related to the way in which linguistic and non-linguistic processing should interact and those related to the representational demands placed on such a model. We also introduce Playpen (Gasser & Colunga 1997) a neural network architecture with these properties. Playpen is a generalization of a continuous Hopfield network in which processing units have phase angles as well as activations and units interact through the attraction and repulsion of their phase angles as well as the usual excitation and inhibition. Phase angles allow the network to bind the features of multiple, simultaneous objects. In addition to units representing features of objects, Playpen also has **micro-relation units** (**MRUs**) representing relation features. Each MRU has two phase angles, one for each of the objects it relates. As in other attractor neural networks, processing consists of pattern completion: certain units in the network are clamped and others are allowed to take on values that are consistent with the clamped values. Though the network is divided into layers of units representing WORDS, CONCEPTS, and VISUAL PROCESSING, none of these layers has any particular priority in processing. Learning in the network makes use of a variant of **contrastive Hebbian learning** (Movellan 1990). Again no units have special priority; any may be treated as input or target with respect to the learning algorithm.

Linguistic and Non-Linguistic Processing

To propose a grounded theory of the learning of words is to argue that the development of word meaning can only be understood in terms of the interaction of linguistic and non-linguistic processing. There are several forms this interaction might take.

1. The way in which the world is construed on particular occasions could have an impact on how language is learned.
2. Specific learned perceptual mechanisms or categories could be prerequisites for the acquisition of specific words or structures.
3. The nature of the human perceptual system could constrain the kinds of meanings that could be learned.
4. Pre-linguistic or extra-linguistic perceptual learning could facilitate the learning of some words more than others.
5. The way in which a language divides up the semantic space in terms of non-linguistic perceptual constructs could affect the order in which words are learned.
6. The regularity implicit in the lexicon of a particular language could lead learners to favor certain patterns of categorization on the part of the speakers of the language.

There is clear evidence for some of these. For example, Regier (Regier 1996) makes a strong case for 3, and Bowerman (Bowerman 1996) and colleagues have data favoring 5. Others, especially those suggested by **linguistic relativism** (6 above), remain more speculative. Our position is that computational modeling may shed light on the possibility of these interactions.

Playpen represents a single model which can accommodate all of these possible effects. Processing can proceed in multiple directions, from vision to concepts to words (as in canonical production), from words to concepts to vision (as in canonical comprehension), from words and a partial visual representation to concepts to a completed visual representation (as in many instances of comprehension), etc. Learning can treat any of the different levels or any combination of the levels as input and target on a given training instance. This non-directionality permits us to investigate a wide variety of interactions between linguistic and non-linguistic categories and processing.

Representational Constraints

What are word meanings? A grounded perspective on this question ties words to perceptual and motoric categories, about which a lot is known. In particular, research on categorization indicates that **object categories** are organized around rich patterns of correlations among object features (Rosch *et al.* 1976). **Relational categories** are more controversial, but work by Billman and colleagues (Kersten & Billman 1997) indicates that these differ more in degree than in kind from object categories and are also defined in terms of featural correlations.

Associating a meaning with a word is in one sense just a further correlation, one between perceptual/semantic features on the one end and lexical or phonetic/phonological features on the other. Like the categories they refer to, words are either objects (nouns) or relations (prepositions and verbs). Thus a noun together with its meaning represents a cluster of object features, and a preposition together with its meaning represents a cluster of relational features.

At the very least, then, representing and learning word meanings requires the capacity to represent and learn correlations between features at different levels. Many neural network models permit this for features of objects (COLOR, SIZE, etc.), but few have anything to day about the primitive perceptual features of relations. Playpen treats relations as composed of binary micro-relations, implemented as MRUs. In Playpen the learning of relational correlations works much the same way as the learning of object correlations, through a simple Hebbian form of learning.

At least two predictions follow from this approach to relational knowledge and learning.

1. Each relation should start in a relatively context-specific form, tied to particular pairs of objects or particular kinds of objects. Later, as MRUs become associated with one another, the relation becomes more abstract. If this is true, we expect children to have more context-specific relations than adults.

2. A major cause of the difference in ease of learning of two relations should be the difference in the correlational structure behind the relations. For example, *left* and *right* are more difficult than *on* and *under* because while *on* benefits from all of the aspects of the world which correlate with vertical orientation, there is little more than horizontal orientation for *left*.

In sum, a grounded model of the learning of word meaning relates linguistic and non-linguistic categories and processes. It should allow for the variety of possible interactions between these two levels. Playpen imposes no direction on processing and learning, so it accommodates all such interactions. A grounded model should also build on an account of what form these perceptual categories take, how they take on that form, and how they are associated with words. Playpen solves these problems through the use of two kinds of primitive units and a simple correlational learning algorithm.

References

Bowerman, M. 1996. Learning how to structure space for language: A crosslinguistic perspective. In Bloom, P.; Peterson, M. A.; Nadel, L.; and Garrett, M. F., eds., *Language and Space.* Cambridge, MA: MIT Press. 385–436.

Gasser, M., and Colunga, E. 1997. Playpen: Toward an architecture for modeling the development of spatial cognition. Technical Report 195, Indiana University, Cognitive Science Program, Bloomington, IN.

Kersten, A. W., and Billman, D. 1997. Event category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23(3):658–???

Movellan, J. 1990. Contrastive Hebbian learning in the continuous Hopfield model. In Touretzky, D.; Elman, J.; Sejnowski, T.; and Hinton, G., eds., *Proceedings of the 1990 Connectionist Models Summer School.* San Mateo, CA: Morgan Kaufmann. 10–17.

Regier, T. 1996. *The Human Semantic Potential: Spatial Language and Constrained Connectionism.* Cambridge, MA: MIT Press.

Rosch, E. H.; Mervis, C. B.; Gray, W. D.; Johnson, D. M.; and Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology* 8:382–439.