Grounding the Acquisition of Grammar in Sensorimotor Representations

Tiago V. Maia and Nancy C. Chang

{maia, nchang}@icsi.berkeley.edu
International Computer Science Institute
1947 Center St., Suite 600, Berkeley, CA 94704*

Abstract

Drawing on data from linguistics, developmental psychology and the neurosciences, we present a computational theory of the acquisition of early grammar by infants. Based on the view that language is a mapping between form and meaning, we propose that a theory of language acquisition must be tightly integrated with a theory of the infant's prelinguistic representations. Namely, the infant's task is to learn how to map the linguistic form in the input to her representations of the corresponding scenes. We have developed a theory of prelinguistic cognition based on i) what is currently known about the architecture of the brain, and ii) the representational requirements for successful (sensorimotor) behavior in the world. We show how such prelinguistic sensorimotor representations can provide the basis for the acquisition of early grammatical forms, and thereby ground language in the world. Importantly, this is true not only at the lexical level, but also at the grammatical level.

Introduction

Approaches to language in the Cognitive Linguistics tradition take linguistic units to be pairings between form and meaning (e.g. Langacker 1991). Typically, elements of form encompass properties of the speech or text stream, such as the identity or ordering of particular segments, while the semantic pole is equated with (non-linguistic) conceptual structure, which is taken to be embodied (Lakoff 1987). That is, the meaning of language is grounded in human sensorimotor systems and, loosely speaking, corresponds to the activation of certain neural structures that represent, in a non-linguistic way, the scene or content associated with the linguistic form (e.g. Regier 1996, Bailey 1997). In this view, the acquisition of language amounts to learning how elements of form map to elements of meaning. This implies that, among other things, one needs a solid theoretical understanding of the prelinguistic structures to which linguistic form will map.

We assume that these prelinguistic structures exist because they are useful for behaving in the world, not merely because they pave the way for language. We base this assumption on an evolutionary line of reasoning: language entered the scene very late in evolutionary time, and as such was built on top of older cognitive skills that we share with many of our relatives in the animal kingdom, such as the ability to move, perceive scenes, act on objects, interact with conspecifics, etc. To some extent, ontogeny seems to proceed in the same manner – before acquiring language, the infant goes through an extended period of physical and social interaction with the environment. Piaget (1952) termed this period sensorimotor, arguing that it was characterized exclusively by motor and perceptual interactions with the world, without any kind of representations or conceptual thought. This view, however, has been the subject of intense debate in contemporary developmental psychology (e.g. Mandler 1992).

In AI, the grounding of single words (or, more generally, symbols) in perceptual and, to a lesser extent, motor systems has been a very active area of research over the last decade (e.g. Harnad 1990, Regier 1996, Bailey 1997, Bailey et al. 1998, Steels and Kaplan 1999). It seems generally accepted that understanding the relation between single words and their underlying sensorimotor representations is fundamental for any model of language acquisition. Importantly, a growing body of evidence from the neurosciences shows that it might also be extremely relevant for language understanding. For example, Pulvermüller (1999) shows that when people hear words typically associated with visual input (e.g. 'bird'), certain areas of the visual cortex become active. Similarly, upon hearing a word typically associated with both visual input and motor actions (e.g. 'hammer'), areas of both the visual and motor cortices become active. This suggests that sensorimotor representations might be directly involved in language understanding. Narayanan (1997) has built a computational model that shows how such sensorimotor representations can also support complex inferences in narratives about abstract domains, through metaphorical mappings. A smaller amount of work in AI has also addressed the grounded acquisition of grammar (e.g. Steels 1997, Oates, Eyler-Walker, and Cohen 1999); this is the problem we address in this paper.

The main defining characteristic of our work is that we explicitly focus on the relation between the acquisition of grammar and prelinguistic structures, basing our account of the latter as much as possible on what is known about the brain. We claim that the prelinguistic structures that enable the child to successfully behave in the world have many characteristics found in language (e.g. relational roles), and that the organization of language partly reflects the organization of such structures. Hence, on our view, the acquisition of grammar is not based merely on

* The first author is also with the Dept. of Computer Science and Engineering, SUNY Buffalo, and the second author with the Electrical Engineering and Computer Sciences Dept., UC Berkeley.

distributional analyses of the linguistic input, but is instead very closely tied to the grounded meaning structures to which linguistic form attaches. We also take very seriously the goal of having a model that is compatible with what is known about child language acquisition.

Prelinguistic Representations

Consider a (pre-linguistic) child observing a scene in which her sister is pushing her brother. What kinds of representations must be involved in the child's understanding of this scene, and how might they be implemented in the brain?

First, it must be assumed that the child has representations for her sister and brother that are independent of this particular scene. These are necessary because the child must be able to recognize her sister and brother independently of the particular actions or perceptual environments in which they appear. We know from the neurosciences that these representations correspond to particular patterns of activity in specific sets of neurons.

We also know that it is likely that there is circuitry in the brain specialized for the recognition of humans. The recognition of conspecifics is evolutionarily very important, and mechanisms specialized for this task have been found in several species (e.g. Sherman, Reeve, and Pfennig 1997). Moreover, neurons that respond selectively to faces have been found in the monkey temporal cortex (e.g. Perret, Rolls, and Caan 1982), and there is ample neuropsychological evidence for double dissociations between the recognition of objects and the recognition of faces, in humans (see for example the prolific literature on prosopagnosia)[1].

A little less intuitively, it seems that the child also needs an independent representation for the pushing, which allows her to understand a pushing action regardless of who is doing the pushing or who or what is being pushed. This claim, while perhaps not immediately obvious, is strongly supported by the discovery of neurons in the ventral premotor cortex (area F5) of the monkey's brain that fire whenever the monkey is either executing a given action (e.g. grasping a raisin), or seeing someone else perform the same action (Gallese et al. 1996). Most relevant for our purposes, these so-called 'mirror neurons', while specific for a particular action, are independent of the agent and the object of the action. For example, a particular neuron selective for grasping might fire regardless of whether it is the monkey, the experimenter or someone else who is doing the grasping, and whether a raisin, a pen or another small object is being grasped. There is also an increasing body of evidence for the existence of mirror neurons in humans (e.g. Rizzolatti et al. 1996). All of this points to the conclusion that the pushing action is represented independently of who is doing the pushing and who or what is being pushed.

Given these considerations, the child's representation of the scene in question must include, at a minimum, the activation of the sets of neurons that represent the sister, the brother and the pushing action. This, however, is not enough. This pattern of activation per se does not signal whether the sister is pushing the brother or the other way around. Therefore, the brain must have some way of indicating that it is the sister who is doing the pushing and the brother who is being pushed.

Now note that this overall conceptual structure, which, ignoring issues of linguistic aspect, corresponds to the meaning of the active transitive sentence 'Sister is pushing Brother', must already exist before language. We can conclude that the machinery for representing entities and actions, and enforcing the necessary bindings between them, taking into account the respective roles, must be pre-linguistic. It is then plausible to assume that, associated with each action, there is a set of roles that must be filled. For example, associated with a pushing action there is always someone who is doing the pushing and something or someone that is being pushed. One could therefore expect a prelinguistic representation of the scene we have been considering as in Figure 1[2].

On the other hand, the fact that, for example, all physical actions are represented in the same area of the brain, would seem to facilitate the formation of a category Physical action[3]. After all, all physical actions will have similar local neural architectures, and similar connections to other areas of the brain. Hence, we may suppose that there are biologically natural categories, of which Physical action is but an example. Other categories one might expect, based on what is known about the architecture of the brain, would include Spatial location, which would be represented in the "where" pathways of the visual system (Ungerleider and Mishkin 1982), Physical object, which would be represented in the "what" pathways of the visual system (Ungerleider and Mishkin 1982), and Human, as we saw above.

[1] We should probably distinguish between recognition at the categorical and individual levels (i.e. recognizing that something is a person versus recognizing that it is a particular person, e.g. John Smith). But it seems likely that both mechanisms are in place and even share certain features.

[2] It is important to note that this figure is not meant to be an accurate diagram of neural circuitry: we have two labeled connections (indicating the roles), and obviously in the brain there are no labeled connections. The roles must therefore be represented via additional sets of neurons. This is the approach in Shastri (1999), where the binding between a role (e.g. PUSHER) and a particular entity (e.g. *Sister*) is achieved by the synchronous firing of the sets of neurons representing the two.

[3] In this paper, we adopt the following notational conventions: categories are indicated by a regular font type (e.g. Physical action), instances (i.e. individuals) by italics (e.g. *Sister*), roles by small caps (e.g. PUSHER), and names of 'relational frames' (i.e. frames with role structure, which relate several independent entities) by bold (e.g. **Push**).

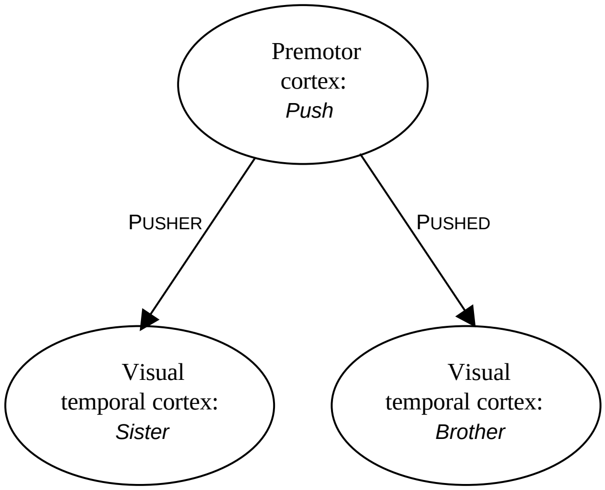

Figure 1 - Prelinguistic representation of a scene in which the infant's sister is pushing the infant's brother.

For example, let us assume that the first time the child notices a pushing action is when she sees her sister pushing her brother. She may represent this scene using an action-specific relational frame for pushing: **Push** [PUSHER: *Sister*. PUSHED: *Brother*] (where we are now resorting to symbolic frames for notational convenience – we have in mind a neural representation in the spirit of structured connectionism (Feldman 1989), such as that in Figure 1, with the roles represented by additional nodes). In neural terms, this pattern of activation will lead to the strengthening of the connections between the push action, the respective roles and their fillers, in a Hebbian fashion. In this way, the **Push** frame above comes to represent the child's (implicit) knowledge of push, stored as a synaptic connectivity pattern[4]. Now imagine that the child later sees her mother pushing her sister. This may be represented symbolically by the frame: **Push** [PUSHER: *Mother*. PUSHED: *Sister*]. The corresponding pattern of activation will change the stored knowledge about push, via Hebbian learning in a cell ensemble that partially overlaps the traces left by the previous instance. Now, because *Mother* and *Sister* are both Human (as are both *Sister* and *Brother*), they share several features and are represented in similar circuits in the brain. It is therefore plausible to assume that the synaptic pattern for push will tend to be generalized towards **Push** [PUSHER: Human. PUSHED: Human] (although it might take several examples, with different instances of Human, for such a generalized frame to be formed). In this fashion, the several action-specific frames will start to make use of general categories.

Through a different process, the action-specific frames themselves may be generalized towards frames that encompass several actions. For example, assume that through the process described in the previous paragraph, the following frames have been formed: **Push** [PUSHER: Human. PUSHED: Human or Physical Object] and **Kick** [KICKER: Human. KICKED: Human or Physical Object]. Now, the symbolic notation used here may mislead one into believing that the roles of these frames are unrelated – there is no apparent similarity between the roles of pusher and kicker. Note, however, that this is an artifact of the notation, which was adopted merely for the sake of expository simplicity. In reality, there are striking similarities between a pusher and a kicker – not only are they both Human, they also both initiate bodily motion aimed at the object of the action, and then, through direct physical contact, exert a force on that object. In a similar vein, there are also obvious image-schematic similarities between pushed and kicked objects (or persons). These similarities between the roles in the two frames may strongly invite the child to generalize across them. More generally, the situation in which a human is physically acting on an object or another human is so ubiquitous that the child will likely form an abstract relational frame of the form: **Directed action** [ACTION: Physical action. ACTOR: Human. PATIENT: Physical object or Human].

Once such abstract representations are available, they will get instantiated with specific values when the child is processing a particular scene. For example, if the child already has the **Directed action** frame, her representation of her sister pushing her brother will be: **Directed action** [ACTION: *Push*. ACTOR: *Sister*. PATIENT: *Brother*]. Note that the relation between the action and its participants may now be captured by the abstract frame, instead of (or possibly in addition to) the action-specific **Push** frame.

In summary, the child might i) start with action-specific frames, ii) later form categories that are biologically natural (e.g. Human or Spatial Location), and, finally, iii) notice relational regularities among these categories, thus creating abstract relational frames. Because there is a small number of biologically natural categories, which get combined in particular ways (e.g. a Human performing a Physical action on a Physical object) there will also be a relatively small number of abstract relational frames. Crucially, we propose that these frames form the basis for grammatical forms that can be found crosslinguistically in child language (what Slobin (1981) calls 'Basic Child Grammar'). For example, as described in the next section, the **Directed action** frame may form the basis for the acquisition of the active transitive construction. We will now turn to evidence from the child language acquisition literature supporting these views.

[4] Hebbian learning typically requires several exemplars for the synaptic connections to be sufficiently strengthened to encode knowledge permanently. Furthermore, the motor schemas for an action (as opposed to an action frame), in motor and premotor cortex, will also depend on several experiences of that action (from an actor, observer or experiencer perspective). For the sake of simplicity, however, we will assume one-shot learning throughout our exposition; nothing we say depends on that.

Language Acquisition

If the child starts by operating with action-specific frames, such as **Push** or **Kick** above, then one expects that, at least in the early stages of language acquisition, grammatical rules will also be action-specific. After all, the child will be mapping the linguistic forms to these representations on an action-by-action basis. Without abstract relational frames that make explicit the similar relational structure of different actions, she will have no particular tendency to make generalizations across actions, or, in linguistic terms, across verbs. This is precisely what has been observed – many early grammatical constructions seem to be verb-specific (Tomasello 1992).

If at a later stage the child is already operating with more abstract relational frames, such as the **Directed action** frame above, then one will expect to observe certain generalization patterns. Namely, whenever the child maps a particular form (e.g. a case marker) to this frame, that form automatically becomes available for all actions covered by the frame. Again, this is what has been observed. Through crosslinguistic studies of language acquisition, Slobin (1981) found that there are "prototypical transitive events" that are especially salient for the child. These correspond to overtly physical actions being performed by an animate being on an object (or, more generally, patient), by means of direct physical contact. Note that this corresponds exactly to our **Directed action** frame. Slobin provides abundant evidence, from studies in the acquisition of several languages, that children consistently make linguistic generalizations across prototypical transitive events that they do not carry over to non-prototypical events. In Russian, for example, the direct object is marked by an inflection, regardless of the type of event; when Russian children first start applying this inflection, however, they use it only to mark the direct objects of verbs involving direct, physical action on things.

The evidence from studies in child language acquisition thus seems to support the developmental timeline sketched in the previous section. Crucially, it also seems to indicate that early grammar reflects prelinguistic structures, which in turn arise from the architecture of the brain.

The idea that the architecture of the brain could be reflected in language is not entirely new. Landau and Jackendoff (1993) show that the existence of divergent "what" and "where" pathways in the visual system is mirrored in language. They observe that the grammatical forms that describe objects and places, count nouns and spatial prepositions respectively, draw on different types of spatial representations. Moreover, the kind of spatial representation each is based on seems to relate to the kinds of representations available in the two visual pathways – the "what" pathway for count nouns and the "where" pathway for spatial prepositions. We take this to be but an example (albeit an excellent one) of the reflection in grammar of the categories that we are biologically predisposed to form. To this example, we might add others, such as the major division in most (if not all) languages between nouns (originating in the representation of objects in the "what" pathways) and verbs (originating in the representation of physical actions in premotor cortex); the active transitive, as just described; the grammaticization of biological gender, which builds on the specialized brain circuitry for determining sex; the grammaticization of number (in English, with the plural), that builds on primitive brain mechanisms for counting; etc.[5]

The following rather provocative (though still conjectural) picture of the organization of language then begins to emerge: different specialized circuits in the brain will tend to give rise to different conceptual categories; these in turn will tend to form the prototypes of grammatical categories. For example, the "what" pathway, the "where" pathway, and the premotor cortex will give rise to conceptual categories that are at the core of the grammatical categories of nouns, spatial prepositions and verbs, respectively. The instances of these grammatical categories (often lexical items, but sometimes also affixes, such as the English plural -s) represent specific distinctions within the corresponding circuits. For example, 'push' and 'grasp', two different verbs, would both be represented in premotor cortex, but with different activation patterns. Our conjecture is that this broad picture holds for the initial stages of both the development and the evolution of language.

Clearly, this is not entirely true for adult language, since it is well known that there isn't a simple, one-to-one correspondence between syntactic and semantic categories. We believe that the route to the adult syntactic system proceeds through a process similar to Schlesinger's (1988) 'semantic assimilation'. Briefly, similarities in syntactic form in the input language lead the child to generalize her syntactic categories to match those that she encounters in the input. This is what turns, for example, 'see' or 'remember' into verbs. Even though they are not clearly physical actions, they become assimilated into the verb category because they occur in the same syntactic situations as physical action verbs such as 'kick' or 'throw'. Importantly, there is also a semantic relation between verbs such as 'see' or 'remember' and physical action verbs – namely, the former are usually accompanied by physical actions. As Schlesinger (1988) puts it, "seeing is not so different from looking, and the latter is action-like; remembering usually goes hand in hand with talking about what is remembered" (pg. 126). This means that the child is able to gradually form a syntactic category, which will exhibit prototypical, radial and graded effects. This accords well with Lakoff's (1987) description of the internal structure of linguistic categories. It is also compatible with Hopper and Thompson's (1980) argument

[5] As pointed out earlier, many of these mechanisms are shared by our relatives in the animal kingdom. See for example Hauser (2000) for descriptions of the capacities in animals for object recognition and representation, the specialized mechanisms for the recognition of sex, their primitive abilities in "number juggling" (Hauser's expression), etc.

that even in adult language some transitive clauses are more fully transitive than others[6].

A similar process could also be at play in the earlier stages of language acquisition, taking the child beyond the verb-specific grammatical constructions. That is, while in this paper we have concentrated mostly on how generalizations in non-linguistic conceptual structure (i.e. the meaning pole) enable generalizations in the form pole, generalizations motivated at least in part by similarities in the syntactic domain may also be pervasive in all stages of language acquisition.

In Chang and Maia (2001) we describe the language acquisition process in more detail. There, we present a computational model of the learning of mappings of relations in form to relations in meaning, based on a probabilistic framework and drawing on the grounded representations described here. We also discuss in more detail the role that similarities in form play in cueing the generalization of grammatical constructions.

Nativism and Empiricism

In taking into account the universal architecture of the human brain, and showing how it affects cognition and language, our approach has a certain dose of nativism. The fact that the gross organization of the brain is similar across individuals, with the same specialized circuits found in the same areas of the brain, seems good evidence that there are genetic determinants at play[7]. Moreover, we know from evolutionary biology that, whenever there is a recurring theme in the environment, evolution tends to endow species with mechanisms specialized to handle it. For example, all humans and many animals need to i) be able to recognize objects, conspecifics and animate beings, ii) have some rudimentary understanding of small numbers, iii) have an understanding of space and the ability to navigate, etc. (Hauser 2000). It then seems reasonable to assume that these mechanisms have been largely built into our brains by evolution.

However, while there are some aspects of cognition that are likely to be genetically determined, many others are not. For example, it is probable that the brain circuitry to recognize objects is at least partly innate; on the other hand, it is hard to imagine that particular object categories (e.g. Telephone) could be. What seems reasonable is to assume that there is specialized circuitry for object recognition, and that there is a (possibly specialized) mechanism to categorize objects by interaction with the environment; the particular categories formed will depend on such interaction.

A similar argument would seem to apply to other major biological circuits that we have discussed in this paper. For example, while the circuitry for action planning, control and recognition is likely to be innate, the particular actions that this circuitry will be able to plan, control and recognize, are most likely acquired, through the composition of basic motor synergies (Bernstein 1967; Bizzi, Mussa-Ivaldi, and Giszter 1991). This suggests a general view in which the major circuits in the brain are biologically determined, but the actual categories and instances within these circuits are often acquired. As we have seen above, our conjecture is that the major circuits give rise to the major grammatical categories in early language development (what Slobin has called ‘Basic Child Grammar’) and evolution.

We also believe that, prelinguistically, the infant is actively categorizing the world, in an unsupervised way. She will likely develop a complex prelinguistic ontology, categorizing together, for example, perceptually similar objects, or objects to which the same sensorimotor schemas can be applied. These considerations suggest that empiricism also has a very important role to play in our approach. In the AI literature, Cohen (2000) and Cohen, Atkin, and Oates (1997) provide excellent examples of an empiricist approach to the acquisition of concepts. We also plan to integrate an empiricist component with our framework, to account for the formation of categories within each of the specialized brain circuits. In short, we believe that the infant forms a prelinguistic ontology with dual foundations – some of the broad categories (e.g. Physical object) are biologically determined, while subcategories of these (e.g. Cup) are acquired.

Before ending the discussion on nativism and empiricism, it is probably worth stressing that, even though we have suggested that there are biologically determined specialized circuits in the brain, we have not found the need to postulate a Chomskyan Language Acquisition Device. Instead, on our view, language acquisition can be accounted for in terms of the relation of language to cognitive systems that precede it, both ontogenetically and phylogenetically. We have seen that this approach has the potential to explain possible linguistic universals (such as the existence of nouns and verbs), not by postulating innate linguistic constraints, but by showing how the structure of language is constrained by the pre-existing structures that ground it.

Acknowledgments

We would like to thank Jerry Feldman for the many discussions and suggestions that greatly helped shape the

[6] It is worth mentioning that the prototypical transitive clause that they identify corresponds to Slobin’s prototypical transitive event, and our **Directed action** frame.

[7] Note that we are not claiming that the specialized circuits in the brain are directly encoded in the genes. There is evidence that initially the cortex is at least partly pluripotential, and that the different afferents to the several cortical areas help determine their function. In general, neural development proceeds through complex interactions between genetically and environmentally determined factors. In any case, even if the functions of the several cortical areas are largely determined by their afferents, the gross patterns of connections of the latter are themselves genetically determined. It is in this broad sense that we speak of ‘genetic determinants at play’.

work reported here, and Ben Bergen and Carter Wendleken for useful comments on this paper. We also extend our gratitude to the other members of the NTL group at ICSI. Opinions and errors remain ours alone. This work was supported by a Praxis XXI Fellowship from the Portuguese “Fundação para a Ciência e a Tecnologia” to the first author, and an IBM Research Fellowship to the second author.

References

Bailey, D.R. 1997. When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs. Ph.D. dissertation, Dept. of Electrical Engineering and Computer Sciences, Univ. of California at Berkeley.

Bailey, D., Chang, N., Feldman, J., Narayanan, S. 1998. Extending Embodied Lexical Development. In *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society (COGSCI-98)*, 84-89.

Bernstein, N.A. 1967. *The Coordination and Regulation of Movement*. New York: Pergamon Press.

Bizzi, E., Mussa-Ivaldi, F.A., Giszter, S. 1991. Computations Underlying the Execution of Movement: A Biological Perspective. *Science* 253:287-291.

Chang, N.C., Maia, T.V. 2001. Grounded Learning of Grammatical Constructions. In *2001 AAAI Spring Symposium on Learning Grounded Representations*.

Cohen, P.R. 2000. Learning Concepts by Interaction. Submitted to the Sixth International Conference on Intelligent Autonomous Systems.

Cohen, P.R., Atkin, M.S., Oates, T. 1997. Neo: Learning Conceptual Knowledge by Sensorimotor Interaction with an Environment. In *Proceedings of the First International Conference on Autonomous Agents*, 170-177.

Feldman, J. 1989. Neural representation of conceptual knowledge. In Nadel, L., Cooper, L.A., Culicover, P., and Harnish, R.M. (editors), *Neural Connections, Mental Computation*. Cambridge, Mass.: MIT Press.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G. 1996. Action recognition in the premotor cortex. *Brain* 119:593-609.

Harnad, S. 1990. The Symbol Grounding Problem. *Physica D* 42:335-346.

Hauser, M.D. 2000. *Wild Minds: What Animals Really Think*. Henry Holt and Company.

Hopper, P.J., Thompson, S. 1980. Transitivity. *Language* 56:251-299.

Lakoff, G. 1987. *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Landau, B., Jackendoff, R. 1993. “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16:217-265.

Langacker, R. W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter.

Mandler, J.M. 1992. How to Build a Baby: II. Conceptual Primitives. *Psychological Review* 99(4): 587-604.

Narayanan, S. 1997. Knowledge-based Action Representations for Metaphor and Aspect (KARMA). Ph.D. dissertation, Dept. of Electrical Engineering and Computer Sciences, Univ. of California at Berkeley.

Oates, T., Eyler-Walker, Z., Cohen, P.R. 1999. Using Syntax to Learn Semantics: An Experiment in Language Acquisition with a Mobile Robot. Technical Report 99-35, Dept. of Computer Science, Univ. of Massachusetts at Amherst.

Perret, D., Rolls, E., and Caan, W. 1982. Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47:329-342.

Piaget, J. 1952. *The Origins of Intelligence in Children*. New York: Basic Books.

Pulvermüller, F. 1999. Words in the brain's language. *Behavioral and Brain Sciences* 22: 253-336.

Regier, T. 1996. *The Human Semantic Potential: Spatial Language and Constrained Connectionism.* A Bradford Book, The MIT Press.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Perani, D. and Fazio, F. 1996. Localization of grasp representations in humans by positron emission tomography: 1. Observation versus execution. *Experimental Brain Research* 111:246-252.

Schlesinger, I.M. 1988. The Origin of Relational Categories. In Levy, Y., Schlesinger, I.M., Braine, M.D.S. (editors), *Categories and Processes in Language Acquisition*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Shastri, L. 1999. Advances in SHRUTI – A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony, *Applied Intelligence* 11: 79-108.

Sherman, P.W., Reeve, H.K., Pfennig, D.W. 1997. Recognition Systems. In Krebs, J.R. and Davies, N.B. (editors), *Behavioural Ecology: An Evolutionary Approach*, 4th Edition, Blackwell Science.

Slobin, D.I. 1981. The Origins of Grammatical Encoding of Events. In Deutsch, W. (editor), *The child's construction of language*, pp. 185-200, London: Academic Press.

Steels, L. 1997. The Origins of Syntax in Visually Grounded Robotic Agents. In Pollack, M. (editor), *Proceedings of the International Joint Conference on Artificial Intelligence 1997 (IJCAI-97)*. Morgan Kauffman Publishers.

Steels, L., Kaplan, F. 1999. Bootstrapping Grounded Word Semantics. In Briscoe, T. (editor), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.

Tomasello, M. 1992. *First verbs: A case study of early grammatical development.* Cambridge University Press.

Ungerleider, L.G., Mishkin, M. 1982. Two Cortical Visual Systems. In Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (editors), *Analysis of Visual Behavior*, The MIT Press.