Natural Semantics for a Mobile Robot

Paul R. Cohen
Department of Computer Science
Lederle Graduate Research Center
University of Massachusetts, Amherst MA 01003
cohen@cs.umass.edu

Carole R. Beal
Psychology Department
Tobin Hall
University of Massachusetts, Amherst MA 01003
cbeal@psych.umass.edu

Abstract

Functionalism is the view that a system (or system component) grasps the meaning of its inputs to the extent that it produces the right outputs. If a system retrieves all and only relevant documents in response to a query, we say it understands the query. If a robot avoids bumping into walls, we say it understands its sensors and its environment. If a chess program beats the world champion, we say it understands the game. One kind of functionalism, *conventional functionalism,* is currently very popular and productive in artificial intelligence and the other cognitive sciences, but it requires humans to specify the meanings of assertions. A second kind of functionalism, *natural semantics* requires computers to *learn* these meanings. This paper discusses the limitations of conventional functionalism and describes some robotics work from our laboratory on natural semantics.

Natural Semantics for a Mobile Robot

Abstract

Functionalism is the view that a system (or system component) grasps the meaning of its inputs to the extent that it produces the right outputs. If a system retrieves all and only relevant documents in response to a query, we say it understands the query. If a robot avoids bumping into walls, we say it understands its sensors and its environment. If a chess program beats the world champion, we say it understands the game. One kind of functionalism, *conventional functionalism,* is currently very popular and productive in artificial intelligence and the other cognitive sciences, but it requires humans to specify the meanings of assertions. A second kind of functionalism, *natural semantics* requires computers to *learn* these meanings. This paper discusses the limitations of conventional functionalism and describes some robotics work from our laboratory on natural semantics.

Conventional Functionalism

A defining characteristic and a great accomplishment of artificial intelligence (AI) is its reduction of semantics to syntax. As Haugeland put it, "If you take care of the syntax, the semantics will take care of itself." An early, celebrated example of syntactic, faux-meaning was Weizenbaum's Eliza system. You could chat with Eliza by typing sentences, to which Eliza would respond with comments that ranged from banal to bizarre to insightful. Eliza did not *understand* the conversation except in the functionalist sense. Consider an Eliza-like dialog: Me: *"You never take out the trash."* Eliza: *"Would you like me to take out the trash?"* Me: *"Yes!"* Eliza: *"Ok."* Needless to say, Eliza doesn't take out the trash and wouldn't even if it were connected to a robot, because it has no clue what "take out the trash" means. All it knows is that the syntactic form, "You never X" can be responded to with, "Would you like me to X?"

Fast forward thirty years, and we find a vast corpus of programs that are much more sophisticated than Eliza but take an *identical* approach to meaning. Many problems can be solved by attending to very narrow aspects of meaning, or to no aspect of meaning at all. For example, some problems in natural language can be solved by attending to statistical properties of word co-occurrences, not word meanings. We have become adept at selecting problems that our programs can solve without attending to meaning; or rather, problems for which syntactic operations such as counting word co-occurrences are proxies for semantic operations such as reasoning about meanings of words.

Moreover, humans are remarkably susceptible to what we might call *semantic illusion*. This is how semantic illusion

works: We provide programs with inputs that are meaningful to us, the programs produce outputs that are meaningful to us; and presto, we think the inputs are meaningful to the machine. An information retrieval system takes a query that is meaningful to us and returns documents that are meaningful to us, and we say it understands the query. It does not. It simply reflects a functionally sufficient sliver of *our* understanding of the query more or less faithfully. All meanings are our meanings, the machine has no *independent* understanding of anything. *We* project *our* meanings onto the symbols manipulated by our programs.

This critical assessment of conventional functionalism should not obscure its many accomplishments. Expert systems, for instance, are masterpieces of conventional functionalism. It doesn't matter that "stiff neck" and "headache" and "meningitis" mean almost nothing to an expert system, as long as the system produces the right output symbol (e.g., meningitis) for a given set of inputs (stiff neck, etc.). Expert systems and many other successful AI systems are essentially scratchpads for us, the users. We assign meanings to the symbols that these systems shift around, and we are often very happy with the results.

Natural Semantics

A natural semantic system is one that *acquires and maintains* meanings for itself. Humans learn meanings of mental states and representations, refining these meanings throughout their lives. Natural semantic systems are not mere scratchpads for meanings assigned by exogenous agents (i.e., programmers), nor must they be told what things mean. You don't require anyone to tell you the meaning of stubbing your toe or finding $100 in your pocket. You know what events and sentences mean because you have learned (or learned to work out) these meanings. Because you are a natural semantic system, we expect you to understand this paragraph and quite literally draw your own conclusions about it. If your conclusions differ from ours, it is not because you have a bug in your program that we need to fix, it is because you maintain a semantic system that overlaps largely but not entirely with ours.

Natural semantic systems are very rare in AI. A case might be made that reinforcement learning produces natural semantic systems, as the meanings of states (i.e., their values) are learned from a reinforcement signal. Sadly, much work in reinforcement learning tries to coerce systems to learn value functions (i.e., meanings of states) that *we* want them to learn. This is accomplished by

fiddling with the reinforcement signal, the state space, and a batch of other factors, until the system becomes a conventional functional system – one that produces what *we* consider to be the correct outputs for a set of inputs. It is terribly ironic that reinforcement learning is capable of producing at least rudimentary natural semantic systems but we use it to produce conventional systems [10].

The Limitations of Conventional Functional Systems

Conventional functionalism is a good way to build systems with a little semantics that do what we want them to, so most AI systems will quite appropriately continue to be based in conventional functionalism. The issue is not whether we should *stop* building such systems, but whether we can actually build conventional functional systems with more than a little semantics, as is required by natural language understanding and other semantically-rich tasks.

Let us also recognize that, for a given, specific task, the semantic content acquired naturally by an intelligent agent can be duplicated or mimicked by a sufficiently motivated knowledge engineer; so the issue is not whether natural semantic systems can think or mean things that conventional functional systems cannot think or mean, but whether there are enough knowledge engineers on the planet to build a conventional functional system that understands as much as we do.

The crux of the argument against conventional functional semantics and for natural semantics is this: To build a conventional functional system, you have to *design* a system whose syntactic operations produce results that you can interpret as meaningful. This design task is *very* expensive, so expensive, in fact, that we have not been able to build conventional functional systems for semantically deep tasks such as natural language understanding. Programming a computer is a paradigmatic example of this design problem. Programs are syntactic machines, and if you want meaningful outputs, you have to write programs whose syntactic operations are meaningful to you. Doug Lenat and John Seely Brown recognized this problem in a paper called "Why AM and Eurisko Appear to Work" [20]. The AM system discovered many concepts in number theory, but when the AM approach was tried in other domains, it didn't work as well. Lenat and Brown concluded that AM worked because syntactic Lisp operations on Lisp representations of mathematical concepts often produced meaningful new mathematical concepts, whereas syntactic Lisp operations on Lisp representations of *other* kinds of concepts rarely produced meaningful new concepts:

> "It was only because of the intimate relationship between Lisp and Mathematics that the mutation operators … turned out to yield a high "hit rate" of viable, useful new math concepts. … Of course we

can never directly mutate the *meaning* of a concept, we can only mutate the structural form of the concept as embedded in some representation scheme. Thus there is never a guarantee that we aren't just mutating some 'implementation detail' that is a consequence of the representation, rather than some genuine part of the concept's intentionality." [20, p.237]

So, Haugeland is correct to say Good Old-Fashioned AI tries to "take care of the syntax, and the semantics will take care of itself," but taking care of the syntax is *very hard*!

Let's be clear about why conventional functional systems are hard to build. All computer-based systems, including natural semantic systems, operate on the "structural form" of concepts, not directly on the meanings of concepts. The problem is maintaining a *correspondence* between the results of syntactic operations and meanings: Too often the results of operations are not meaningful. In conventional functional systems, the system itself cannot check whether its results are meaningful because it doesn't know the meanings of its results – all meanings are exogenous, maintained by humans, and so only humans can check whether the system's conclusions make sense. This checking, and subsequent corrections, are very expensive.[1]

In natural semantic systems, operations on data structures are also syntactic, but the system itself is responsible for maintaining the correspondence between syntactic forms and their meanings, and the system itself checks whether syntactic operations produce meaningful results. Suppose a robot is planning to move to an object. Syntactically, it generates a plan to rotate the left wheel in one direction and the right wheel in the other direction. When it executes this plan, it spins around. The *meaning* of this plan is, among other things, that it does not achieve the robot's goal of moving toward the object. But we do not have to tell the robot this, because the robot maintains the semantics of its plans, itself.

Natural Semantics for a Robot

This is how it is supposed to work – indeed, does work, to a limited extent – in our laboratory. We provide a Pioneer 1 robot with a handful of activities, little programs for moving, turning, opening its gripper, and so on. We also provide the robot with something like curiosity, a drive to explore things it hasn't done recently or doesn't recognize. The robot moves around its environment, collecting

---

[1] Logic is supposed to solve this problem for us by allowing only truth-preserving inferences. All such valid inferences produce meaningful conclusions so nobody need check whether the conclusions are meaningful. This logicist position has many weaknesses, not the least of which is that truth preserving inferences may produce objects that have no denotation in the world, or no interest to us, as Lenat and Brown discovered.

experiences. Experiences are just time series of the sensor vector (which currently contains roughly 40 sensed values sampled every 100 msec.) We run statistical procedures to find common sequences in the time series of the sensor vector [1]. We cluster these sequences, and in the process discover clusters that correspond to activities such as bumping into a wall and grasping a cup. We call these *prototypes* – sensory patterns that correspond to the robot's common activities [2,3]. It is not difficult to turn prototypes into planning operators [4,6]. Then, with a simple planning algorithm, our robot has started to construct more complex activities in service of its goals. Its goals, not ours, as dictated by its curiosity mechanism.

Each of the robot's activities has roles. When *we* observe or describe activities, we assign roles to the participants. For example, *you* are the reader, *this* is what you are reading, *we* wrote it, and the text itself is conveyed by some *medium* such as print. Familiar roles such as actor, action, object, subject, and instrument, may be augmented with more specific roles such as "the object on my left." General or specific, the idea is the same: A scene is described and presumably represented in terms of the roles of its participants – the causal relations that hold among the participants' actions.

Roles are the key to having our robot learn an ontology through interaction with its environment. Just as we may define "chair" in terms of interaction – a chair is something we sit in – so does our robot ontology include objects that fit between the grippers, objects that can be lifted, objects that do not move, objects that are so large that they are not entirely visible when the robot touches them, and so on. Classes of objects follow directly from roles; indeed, for each role played by an object (e.g., being picked up by the robot) there is a class of objects that have played or could play that role. The same point holds for actions which play roles in activities. If the robot could perceive the roles played by objects and actions, it could develop a conceptual system, an ontology, an organization of objects and actions into classes defined by experience.

Much of the meaning of a token, such as a word, or a sonar reading, is in the roles played by the denotation of the token. For example, "table leg" denotes a piece of furniture that often causes trouble for our robot. To the robot, "table leg" evokes experiences of abrupt deceleration and complicated attempts at disengagement. To us, "table leg" means something else, although the robot's meaning is not entirely alien to us. Still, it's a stretch to say we and the robot have a shared understanding of the phrase "table leg." So how does the story so far account for shared meaning and convivial computers. *A token means the same thing to you as it does to me if it denotes something that plays the same roles in your experiences as it does in mine.* To the extent that our robot has similar experiences as we have, the tokens that denote aspects of those experiences will have similar meanings.

How much of this story is fantasy and how much exists today? We have a robot that creates prototypes that correspond to its common experiences [1,2,3]. It can cluster these prototypes and discover, entirely without our supervision, common experiences like moving past a cup or bumping into a wall. It can also recognize words and link them up with experiences [7]. Recently, it turned some of its prototypes into planning operators, and built a plan for a goal of its own choosing [4]. The missing piece of our story concerns roles. The prototypes learned by the robot correspond to the *sensory* experience of doing something, but they are not denoting representations. For example, the robot knows how its sensors react to driving into a wall, but it has no concept of wall, and it does not represent the episode as one entity doing something to another.

The Roles Problem

The roles problem involves getting from sensors to "who did what to whom." It will be clear that the roles problem is an instance of a more general perception problem [16,11]. Perception establishes the connection between moving patches of light and objects in the environment. Perception establishes the roles these objects play in scenes we observe.

Like it or not, we cannot solve the roles problem without giving our robot a perceptual system. However, we needn't "solve the vision problem" as a prerequisite to natural semantics, because a rudimentary perceptual system will suffice. All we need is a way to map from sensory patterns to scene elements. The solution might be as simple as identifying the scene elements by their characteristic color and size. But instead of cobbling together a perceptual system, we want to build one that has the essential characteristics of infant perceptual systems.

According to Bornstein [12], infants are preprogrammed via neural mechanisms with a limited ability to recognize an object as the same thing when encountered under different conditions. At first (i.e., at birth), this means recognizing an object as the same under different lighting, at various distances, and from different orientations. That is, there is a bias to assume that a pattern of perceptual information that serves as a perceptual anchor in the array while other aspects of the environment are changing must be "the same thing". We are currently providing this bias to the robot.

There also appears to be an early-established coordination of perceptual information from different sensory systems [17]. Infants are able to recognize (in the sense of responding to the object as if it was familiar) objects encountered in one sense and re-presented via a different sense (oral; visual; manual exploration). This "cross-modal transfer" implies that perceptual information is being integrated into a central representation of "an

object". By analogy, we use the temporally synchronized information from different systems in the robot (vision; gripper; wheel movement) as the basis for a primitive perceptual recognition system [8].

However, the detection of perceptual identity and similarity is not sufficient for a natural semantics. Conceptual understanding requires more than merely the ability to see that two things *look alike*; it requires an understanding that two things *are alike*. Human toddlers assume, without ever being explicitly told, that caterpillars and emus share the same kinds of "innards" whereas cars have something different inside. Conversely, no human infant would behave in the same manner towards a real cat and an animated stuffed-animal cat, no matter how realistic the toy version might be. The essential difference in meaning is that some things are alive whereas others are not. The next step is therefore to find a way to extract meaning from perception; we propose to implement a process known as *perceptual redescription*.

Mandler [21,22,23] proposes that infants do not merely perceive the world; they select or extract a subset of information from their percepts, at the cost of other information. The retained information is a redescription of the original perceptual event. Mandler argues that what's extracted from the perceptual analysis is an *image schema*, which is a form of representation in which the spatial structure of the represented world is preserved, including the motion of objects in the world and temporal relations, in the representing world. An example of an image schema is PATH: a pattern of motion with a start and an end, and a direction, that can be regular or irregular. CONTAINMENT; AGENCY; and CAUSED MOTION are others. Image schemas can be combined and embedded to form more complex meanings: thus, ANIMAL is an agent that starts on its own, and has an irregular path of motion. Even temporal concepts can be represented in spatial terms: we speak of one thing occuring "before" (in front of) another, or "after" (behind).

Mandler argues that these spatial patterns and relations are the core meanings that underlie all of human semantics, a claim that has considerable support from work on infant concepts. The notion of image schemas also finds support in linguistics and philosophy [15,18,19]. Some empirical evidence that children really do develop image-schema like mental structures has been gathered by [13,14,24,25,26].

It is of course possible that the particular set of image schemas proposed by Mandler are not the "right ones". It is also possible that the entire notion of going from a richly detailed array of perceptual information into a small set of spatial redescriptions is not the right way to build a conceptual system. Yet the advantages of the perceptual redescription theory are that a) it is testable; b) it seems to be the way that human infants build the basis for a natural semantics, and c) it provides a reasonable solution to the greater problem of where meanings originate. As philosophers have noted, there must either be a small set of core primitives from which all of semantics is constructed, or we must be born with all possible meanings pre-encoded. The latter claim is developmentally implausible in the case of human infants, and of little help to the goal of building convivial computers. By contrast, image schemas are developmentally plausible and are specific enough to implement in a robot. Indeed, so general are image schemas that we have built a planner for tactical wargames based on them [9].

Grant us a moment of speculation on the kinds of concepts and meanings that our robot will learn and share with us. Some aspects of the robot's experience are alien to us, but much is common. For example, we move on planar surfaces from one place to another, so does the robot. We see things in the foreground more clearly than things in the background, so does the robot. The optic flow of foreground objects is relatively high for us and for the robot. We can move something by pushing it, so can the robot. When we turn, objects enter and leave our field of view, as they do for the robot. When we bump into something, we sense force, as does the robot. In sum, many physical activities are essentially the same for us and for the robot.

Does this mean we are condemned to discussing only physical activities with our robot? A really convivial computer would understand the mental world, too. Are we limited to discussing pushing and moving with the robot or can we also discuss our busy day, our delightful children, our new research? Can the robot learn such concepts and their meanings? Two pieces of evidence give us hope.

First, infants are born sensorimotor agents not unlike our robot. They don't think, they interact. Therefore, all knowledge is grounded in physical interaction – be it ever so abstract and non-physical, this knowledge is learned by an intelligence that once was purely physical. Second, some linguists (e.g., [18]) believe that the semantic basis for nonphysical concepts is physical concepts . For example, we are developing a *line of argument*, which *leads to* a hopeful conclusion; you *grasp* it as it *takes shape*. The *missing piece* of the argument is the mechanism for *extending* physical concepts into a nonphysical *space*. Linguists say this mechanism is metaphor. All the italicized words in the previous sentences are physical concepts used metaphorically here.

In summary, we have started to build a natural semantic system, a robot that learns concepts and meanings by interacting with its environment. Whereas a functionalist would claim that a robot understands walls if it avoids crashing into them (or solves some other problem set by the functionalist), we claim that the token "wall" means something *to the robot*, something related to the *robot's* experiences with walls (as distinct from experiences that

we programmed the robot to have); and to the extent that we have had similar experiences with walls, we share a common understanding of "wall" with the robot. We are very hopeful that robots will be able to talk to us not only about their physical experiences, but also about mental concepts that are metaphorical extensions of physical concepts.

References

[1] Names removed for blind review. 1999. Identifying Qualitatively Different Experiences: Experiments with a Mobile Robot. To appear at The Sixteenth International Joint Conference on Artificial Intelligence.

[2] Names removed for blind review. 1999. Abstracting from Robot Sensor Data using Hidden Markov Models. To appear at The Sixteenth International Conference on Machine Learning.

[3] Names removed for blind review. 1999. Continuous Categories for a Mobile Robot. To be presented at The Sixteenth National Conference on Artificial Intelligence.

[4] Names removed for blind review. 1998. Learned Models for Continuous Planning. Presented as a poster at Uncertainty '99: The Seventh International Workshop on Artificial Intelligence and Statistics.

[5] Names removed for blind review. 1998. Applicability of Reinforcement Learning. In The Methodology of Applying Machine Learning: Problem Definition, Task Decomposition and Technique Selection Workshop, ICML-98. Pp. 37-43.

[6] Names removed for blind review. 1998. Learning What is Relevant to the Effects of Actions for a Mobile Robot. In Proceedings of the Second International Conference on Autonomous Agents. Pp. 247-253

[7]. Names removed for blind review 1999. Using Syntax to Learn Semantics: An Experiment in Language Acquisition with a Mobile Robot. Submitted to The Twenty-First Annual Meetings of the Cognitive Science Society

[8] Names removed for blind review. 1997. NEO: Learning Conceptual Knowledge by Sensorimotor Interaction with an Environment. In *Proceedings of the First International Conference on Autonomous Agents*, pp. 170-177.

[9] Names removed for blind review. 1999. Capture the Flag: Military simulation meets computer games. Working notes of the AAAI Spring Symposium on AI and Computer Games.

[10] Names removed for blind review. 1998. The Applicability of Reinforcement Learning.

[11] R.Baillargeon. How do infants learn about the physical world? Current Directions in *Psychological Science*, 3:133--140, 1994.

[12] Bornstein, M. H. (1984). A descriptive taxonomy of psychological categories used by infants. In C. Sophian (Ed.), Origins of cognitive skills, pp. 313-338. Hillsdale NJ: Erlbaum.

[13] G. J. Gibbs and H. L. Colston. The cognitive psychological reality of image schemas and their transforms. *Cognitive Linguistics*, 6--4:347--378, 1995.

[14] E. J. Gibson. Detection of the traversability of surfaces by crawling and walking infants. *JEXP: Human Perception and Performance*, 13:533--544, 1987.

[15] Mark Johnson. The Body in the Mind. University of Chicago Press, 1987.

[16] S. S. Jones and L. B. Smith. The place of perception in children's concepts. *Cognitive Development,* 8:113--139, 1993.

[17] K. L. Kaye and T. G. R. Bower. Learning and intermodal transfer of information in newborns. *Psychological Science*, 5:286--288, 1994.

[18] Lakoff, G. 1984. Women, Fire and Dangerous Things. University of Chicago Press.

[19] Ronald Langacker. Foundations of Cognitive Grammar, Vol. 1. Stanford University Press, 1987.

[20] Douglas B. Lenat and John S. Brown. Why AM and EURISKO Appear to Work, AAAI83. Pp. 236-240.

[21] Jean M. Mandler. How to build a baby: On the development of an accessible representational system. *Cognitive Development*, 3:113--136, 1988.

[22] Jean M. Mandler. How to build a baby: II. Conceptual primitives. *Psychological Review*, 99(4):587--604, 1992.

[23] Mandler, J. M. (1998). Representation. In W. Damon (Series Editor), Handbook of child psychology, Volume 2: Cognition, perception and language (D. Kuhn & R. S. Siegler, Volume Editors) (pp. 255-308).

[24] E. S. Spelke, K. Breinlinger, J. Macomber, and K. Jacobson. Origins of knowledge.. *Psychological Review*, 99:605--632, 1992.

[25] Elizabeth S. Spelke. The origins of physical knowledge. In L. Weiskrantz, editor, Thought Without Language. Oxford University Press (Clarendon), 1988.

[26] E. Thelen and L. Smith. A Dynamic Systems Approach to the Development of Cognition and Action. MIT Press, Cambridge, MA, 1994.

[27] Weizenbaum, J. 1966. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9(1):36-44.