A Dynamical Basis for the Semantic Content of Verbs

Paul R. Cohen and Tim Oates
Experimental Knowledge Systems Laboratory
Department of Computer Science
Box 34610 LGRC
University of Massachusetts
Amherst, MA 01003-4610
{cohen, oates}@cs.umass.edu

Much of what we know and say refers to the dynamics of our world. Here we include our physical world, as well as our mental world, the world of social interactions, and other not-entirely-physical environments. We have a large class of linguistic objects – verbs – devoted entirely to expressing dynamics. Subtle differences in the meanings of verbs, which linguists call "manner", are also often dynamical (Talmy 1985). For instance, the difference between "nudge" and "shove" is partly a matter of mass, movement, and energy transfer from one body to another; and partly a matter of intention.

In this paper we present a dynamical representation sufficiently general and expressive to capture both the meaning of individual verbs and the more abstract commonality of sets of semantically related verbs. Our purpose is to develop a complete theory of language acquisition that can be implemented on a physical platform, such as a mobile robot. The representation of the semantics of a word or set of words is constructed incrementally based on what is happening in the agent's environment when the word or a member of the set of words is heard.

Dynamical representations have a number of advantages for representing and acquiring the semantics of verbs. They are grounded in the sense that one can attach sensors to a corpus of dynamical concepts and have the corpus recognize concepts from sensed movement (Rosenstein *et al.* 1997). Dynamical representations of physical interactions are easily learned from observations of dynamics (Rosenstein *et al.* 1997), this is true also of dynamical representations of linguistic constructs (Regier 1995; Elman 1995). They are compact in the sense that a single representation can describe dozens of related concepts. They make explicit the manner of movement and thus make fine distinctions between word meanings. The strongest reason to consider dynamics as a foundation for semantics, we think, is that the knowledge of the youngest humans - neonates and infants - is produced by interacting physically with the world. Neonates are capable of movement, but nobody credits them with conceptual thought. Concepts must therefore result from neonatal and infant experience, which is primarily sensorimotor experience.

Consider the interaction between the two agents shown below in Figure 1. That interaction decomposes quite naturally into three distinct phases: before, during and after contact. Prior to contact, one of the agents alters its path, apparently to intercept the other. We could call this segment of the interaction "chasing". The duration of contact is brief, yet forceful; the chaser "hits" the chasee. Once contact is broken the agents continue moving in close spatial proximity, as if one of them is now "following" the other. The semantics of a remarkably large number of verbs can be cleanly captured with this before/during/after decomposition.

We want a representation that a situated agent watching or engaging in the above interaction could construct given readily observable features. Cartesian coordinates (x and y locations) are not suitable features because, intuitively, verb roots never encode actual spatial locations. Therefore, we plot interactions in the two-dimensional spaces shown below in Figure 2 and Figure 3. We call each of these spaces a *map*. Maps are constructed by watching interactions unfold in cartesian space and tracing the corresponding path in the map. The axes of the before and after maps, which should capture features of the motion of two agents relative to each other, are the same: the distance between the two agents, $D(A, B)$, and the relative velocity of the two agents, VR. As shown in Figure 2, VR is calculated by considering only the component of velocity that is directed toward (or away from) the other object. During contact we plot the relative energy expended, $E(A, B)$, and the distance from the point of contact, $D(AB, PC)$.

Figure 4 shows several trajectories through our tripartite decomposition of interactions between agents.

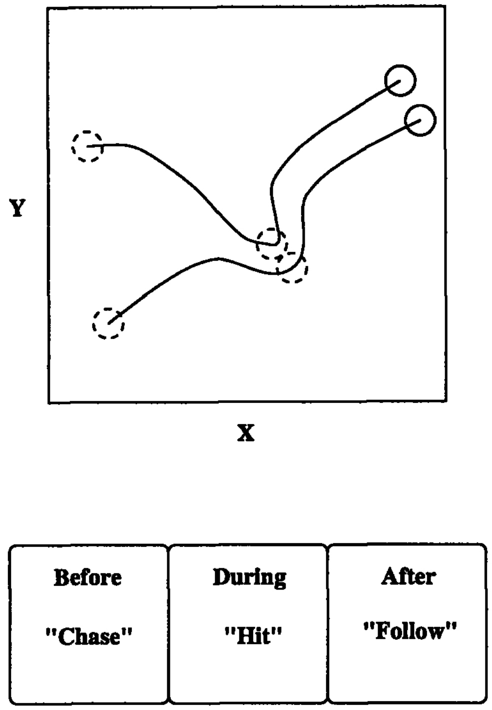

Figure 1: An interaction between two agents that decomposes into three segments: before, during and after contact.

The examples described below are labeled with three characters which correspond to the paths in the before, during and after maps respectively.

- **aaa:** A approaches B, touches it, and remains in contact with it. A gently touches B with no net transfer of energy between them. Relative velocity is inherently ambiguous: We know A and B have equal velocities in the after phase, but we don't know whether this velocity is zero.
- **ada:** A approaches B, makes contact, then gradually increases the energy it transfers to B, maintains a level of energy transfer, then ramps down. A and B remain in contact in the after phase. A pushes B.
- **adb, acb:** A approaches B, makes contact, and gradually (d) or rapidly (c) increases the energy it transfers to B. In the after phase, B moves a little ahead of A. Initially its velocity increases relative to A's then decreases. Depending on the rate of energy transfer, the amount transferred, and the distance B moves in the after phase, this is kick, nudge, shove, propel, and so on. The movement in the before phase is inherently ambiguous: We don't know whether A is moving toward B, B is moving toward A, or both. Similarly, the increasing distance between A and B in the after phase might occur because A stops moving (or slows down) but B continues, or because B stops and A is recoiled, or a combination of effects. Thus, **acb** represents A bounces off B as well as kick, shove, and so on. Similarly, **acb** represents symmetric repulsion, where A and B approach each other, make contact, then bounce away from each other.

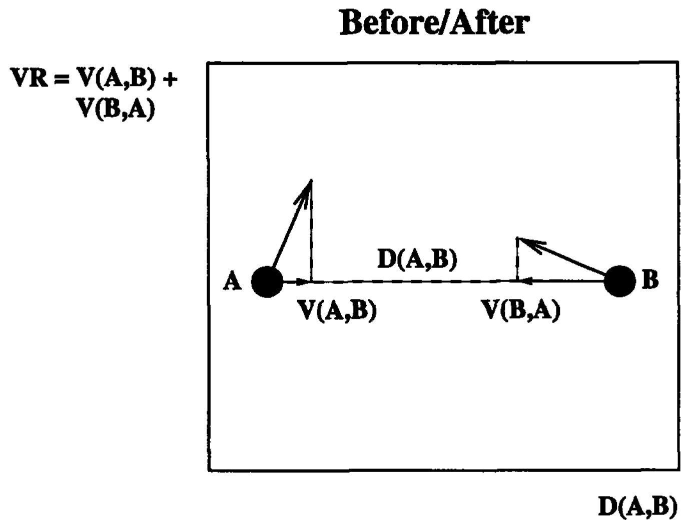

Figure 2: The axes of the before and after maps are relative velocity and distance.

Maps can also be used to find abstract, dynamical characterizations of the semantics of sets of verbs. This is important for at least two reasons. First, individual languages differ in which semantic features are lexicalized in verbs (Talmy 1985). Identifying those features can facilitate acquisition of the semantics of additional verbs by focusing the learner's attention on particular aspects of the environment. Second, there is evidence that simple dynamical semantic features of verbs, such as whether they involve motion or contact, serve to divide verbs into classes that determine allowable argument structures (Pinker 1989). That is, semantic features of classes of verbs, grounded in dynamics, might be crucial to explaining syntactic phenomena. We performed experiments in which simple distributional clustering techniques were used to construct a hierarchy of verbs that behave the same syntactically, often yielding semantically coherent classes (Pereira, Tishby, & Lee 1993; Redington, Chater, & Finch 1993). Each node in the hierarchy was paired with a map that was filled in when any of the words beneath that node occurred. The result is a map that captures the dynamical features that are common to all of those words.

Given that verbs denote activities, it seems clear

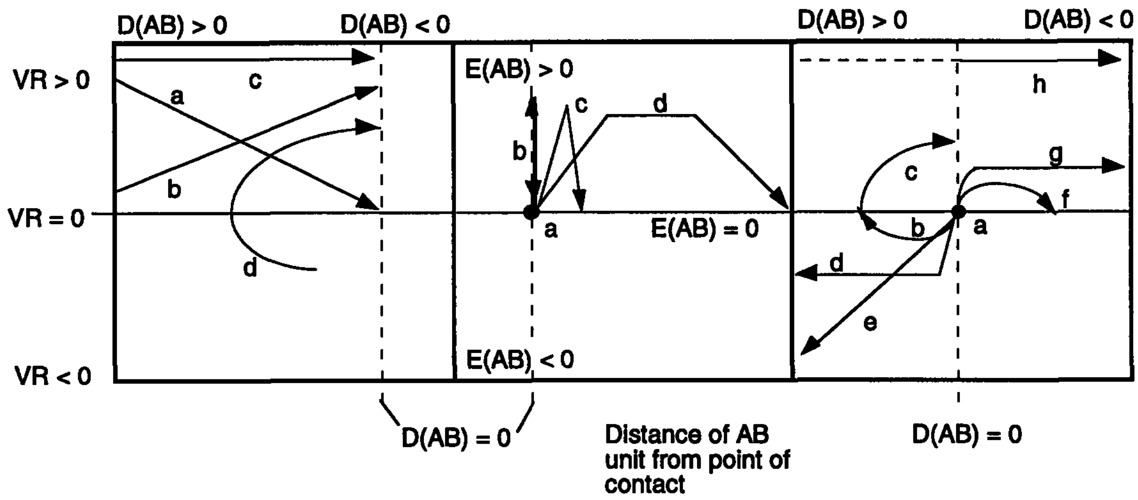

Figure 4: Examples of meaningful paths through maps.

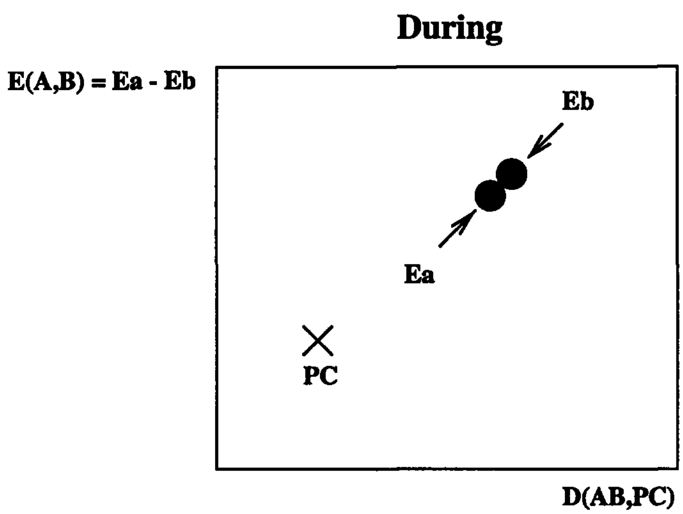

Figure 3: The axes of the during map are relative energy expenditure and distance from point of contact.

that dynamical representations are a good choice. What may be less clear is that dynamical representations are also appropriate for the semantics of words denoting objects. In the interactionist view, which is attributed to Lakoff and Johnson (Lakoff 1984; Lakoff & Johnson 1980) and to which we subscribe, category distinctions are based on activity. Objects and classes of objects are differentiated by how we interact with them. For example, one could define spoons, and thus the semantics of the word "spoon", in volumetric terms, or in terms of the materials from which spoons are fabricated. But that's not how we think of spoons unless we're trying to design or fabricate spoons, so even in this case the definition is tied to activity. So the concept of spoon is really a representation of the activities spoons are involved in, and the meaning of this concept is essentially predictive: What it means to be a spoon is just what happens to spoons in various activities.

Due to the importance of dynamics in capturing the meaning of words denoting activities and objects and the role of dynamical features in syntax, we are optimistic that our focus on dynamics and the representation that we presented above can serve as the foundation for a comprehensive theory of language acquisition.

References

Elman, J. 1995. Language as a dynamical system. In Port, R. F., and van Gelder, T., eds., *Mind as Motion*. MIT Press. 196–225.

Lakoff, G., and Johnson, M. 1980. *Metaphors We Live By*. University of Chicago Press.

Lakoff, G. 1984. *Women, Fire, and Dangerous Things*. University of Chicago Press.

Pereira, F. C. N.; Tishby, N. Z.; and Lee, L. 1993. Distributional clustering of english words. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, 183–190.

Pinker, S. 1989. *Learnability and Cognition: the Acquisition of Argument Structure*. MIT Press.

Redington, F. M.; Chater, N.; and Finch, S. 1993. Distributional information and the acquisition of linguistic categories: A statistical approach. In *Proceed-*

ings of the Fifteenth Annual Conference of the Cognitive Science Society.

Regier, T. 1995. A model of the human capacity for categorizing spatial relationships. *Cognitive Linguistics* 6(1):63–88.

Rosenstein, M.; Cohen, P. R.; Schmill, M. D.; and Atkin, M. S. 1997. Action representation, prediction and concepts. University of Massachusetts Computer Science Department Technical Report 97-31, also presented at the 1997 AAAI Workshop on Robots, Softbots, Immobots: Theories of Action, Planning and Control.

Talmy, L. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Shopen, T., ed., *Language Typology and Syntactic Descriptions III.* Cambridge University Press. 57–149.