THESIS PROPOSAL

# Getting a Grip: Body-Based Acquisition of Verb Semantics for Hand Actions

by

David R. Bailey
Computer Science Division
University of California at Berkeley

Qualification Exam Committee:
Professor Robert Wilensky, Computer Science, Chair
Professor Jerome Feldman, Computer Science, Advisor
Professor George Lakoff, Linguistics
Professor Dan Slobin, Psychology

November 9, 1994

## Abstract

How do people learn to describe simple actions, such as lifting a mug or pushing a box? By way of a verb learning task, we bring together constraints from linguistics, psychology, neurobiology, and computer science in an attempt to discover some underlying universals in the cognitive representation of simple intentional activities. Our model must learn verbs for hand actions from any natural language in such a way that it can both describe novel actions, and generate appropriate actions given a verbal command. Our current model utilizes a state-machine representation for controlling activity called "schemas". Verb learning is aided by the extraction of salient features from executing schemas; in turn, these features are used to guide schema execution. If this model proves successful, it should should support the learning of new hand activities solely by linguistic instruction. Furthermore, it is hoped that our representations can provide the underpinnings of metaphorical understanding of more abstract concepts.

# Contents

# Chapter 1

# Introduction

## 1.1 Task Definition

How do we learn to describe what we do?

Our task is to model children's learning of the semantics of verbs referring to motor activities. In order to provide a focus we limit the domain to hand actions; indeed, we will consider only those actions which can be performed with a single hand upon solid objects on a tabletop. These verbs will involve object movement, obtaining and releasing possession, and object manipulation. The model must learn the verbs from a reasonable number of examples, for any natural language. For each language, we are interested in a small set of words which cover this domain, which appear early in children's vocabularies, and which make qualitative distinctions. For English these verbs include:

| | | | | | |
|---|---|---|---|---|---|
| push | shove | get | put | touch | pinch |
| pull | slide | pick up | let go | feel | squeeze |
| move | lift | grab | drop | hit | tear |
| yank | lower | catch | throw | tap | rub |
| drag | | hold | | | scrub |

More precisely, we want to implement a system which learns the verbs of any single natural language from a set of training examples each containing:

- an action
- a verb labelling the action.

The resulting system must be able to:

- label novel actions
- carry out appropriate actions when given a verb as a command.

It is crucial to note that we are modelling children's ability to label *their own* behavior, not that of others. Therefore, the representation of actions will include—indeed spotlight—intentions and motor commands, although perceptual feedback will be included as necessary. We use the term **schemas** for our representation of the control of activity. Also note that in order to obey a verbal "command", the system must combine semantic constraints with constraints provided by the world. Certain verbs may not be executable in certain world states, and the system must report this.

This task description extends the original $L_0$ domain (Feldman *et al.* 1990; Feldman *et al.* 1994) but maintains its spirit. In particular, this task is closely analogous to Regier's thesis (Regier 1992) which modelled the semantics of spatial relation terms based on visual representations. That task and ours both involve learning single lexemes from a set of positive-only examples of the form <lexeme, semantics>. In both models, the semantics are *bodily grounded*, but are represented at a relatively high level motivated by what is known about the lower levels. Furthermore, both problem statements are designed to uncover universals, by virtue of the requirement to learn any natural language which prevents us from "hardwiring" too much into the model. Lastly, the tasks share the success criteria of judging the model's ability to label novel situations.

However, the focus upon motor commands is a departure from previous $L_0$ work. Just as original $L_0$ had to justify its proposed high-level representation of incoming (perceptual) information, we must also justify our high-level representation of outgoing (motor-control) information. Also, our verb-learning task carries a new criteria for success, namely the ability to carry out actions based on a verbal command. Thus, we must build *active* representations, not just a recognizer. [1]

Methodological details including our training strategy, evaluation criteria, and our use of a body simulator/animator, may be found in Appendix A.

### 1.1.1 What's Left Out

Since we consider learning verbs in isolation from the rest of language, the current task does not address grammatical issues at all. We assume that other mechanisms have determined that a new word is indeed a verb and that it refers to the activity with which it is paired. We ignore inflections, assuming other processes factor out these variations when learning the semantics of verb roots. We do not model verb valence. While this was not a serious shortcoming in Regier's model—where all spatial terms required a Trajector and Landmark—it is more problematic in this project since our verbs vary in valence. We can partially compensate by designing our schema representation to provide details on exactly those thematic roles which are relevant to the activity. If there are substantial interactions between any of these other processes and the semantics acquisition process, we will miss them.

Our intent to learn these verbs strictly in terms of percepts and motor control—excluding other properties such as mood, higher-level goals, etc.—may lead to an unnecessarily complicated view of their semantics. While we acknowledge that we will not be modelling the full semantics of our verb

[1] Some effort has been made to extend Regier's model to be active, i.e., to transform a spatial term paired with a trajector and landmark into a prototypical "mental image". My earlier "Inference through Imagery" project (Bailey 1992) used the technique of backprop through the inputs (Thrun 1992) to generate a prototypical image (at the theta-node level) for a given relation. Simultaneous constraint satisfaction can yield an image with multiple objects, from which novel relational queries can be answered by the original recognizer. Major limitations included the lack of a serious model for the binding mechanism, the inability to handle discrete features such as the bitmap-based features, slowness of the settling process, and vulnerability to local minima. Most importantly, I find it implausible for imagery of a single relation to involve a settling process of this sort (although imagery of multiple relations might well involve settling).

set, we believe that motor-control is a central of the full semantics, and that it is also relatively well-understood, and thus studying the mapping from motor-control to verbs is an interesting and worthwhile undertaking.

### 1.1.2 Possible Extensions

There are several intriguing ways to extend the basic task. An obvious path is to gradually enrich our schema representation so that the model can handle more subtle word distinctions or more complex activities. This might include an "interrupt-handler" mechanism, or even a simple planner, to handle unexpected contingencies, allowing learning of words such as "stop", "whoops", "uh-oh", "try", etc.

Another incremental change would be to extend the domain to include some high-level properties (for example, extending objects to include "types" such as "tool" or "toy"), or by introducing jointed segments (facilitating *open* and *close*), or by modelling liquids or other kinds of deformations (facilitating *spill* or *crumple*).

We could add to the model a component that builds schemas for new tasks, solely from a linguistic description of the new task in terms of already-learned verbs. This would be a step towards demonstrating that our representations are adequate for reasoning, since such a model would have to address issues such as protection intervals, maximizing concurrency, combining trajectory constraints, and understanding "glue" words such as "then" and "while".

Moving in the other direction, we might go back and consider how the schema set (and other currently hardwired portions of the model) might be reorganized during language learning, in an attempt to explain or predict any relevant "Whorf effects"—differences in cognitive skills resulting from language differences. We might also consider how schemas can be learned in the first place, via operant conditioning, capitalizing on the perceptual and behavioral biases known to be hardwired in humans. The relative simplicity of our current schema representation may permit a Hidden-Markov-Model-induction solution using the algorithms in (Stolcke 1994); initial experiments showed this approach to be feasible.

## 1.2 Motivation—"It's the Body, Stupid!"

Key to this project is our belief that many conceptual domains, and especially a domain such as intentional activity, are fundamentally *bodily grounded* (Johnson 1987). That is, our concepts are rely more upon the particulars of the human body and the way it interacts with the world, including perceptual- and motor-system characteristics, brain characteristics, emotions, biomechanics, the autonomic system, and so on, than upon "objective", abstract representations of the external world.

A canonical example of the study of bodily-grounded cognition is the color-terms work of Berlin & Kay (1969) (also see (Shepard 1992)), which showed that while languages differ in the number of color terms employed, there is a priority ranking by which colors will be chosen. The prototypes for these colors are uniform across languages, and can be predicted from the physiology of the visual system. We hypothesize that the semantics of hand verbs in different languages similarly arise from varying combinations of a small set of primitives which in turn derive from the parameterizations of motor system synergies. Another landmark study in bodily-grounded cognition is the work of Rosch on basic-level categories (Rosch *et al.* 1976), which showed that object hierarchies exhibit one special level of abstraction at which most reasoning is performed, and this level turns out to be the highest level which permits mental imagery. That work concentrated on nouns; we apply this lesson to verbs by choosing to investigate reasoning about activity via a representation that supports mental simulation of actions.

There is much evidence (Lakoff 1987) that such bodily-grounded semantics may underlie not only the "literal" interpretations of these verbs, but also their metaphorical meanings. Metaphorical and other mappings from the bodily domain to more abstract domains may in fact explain a large chunk of human reasoning, lending added importance to the goal of understanding body-based representations. Talmy (1988) presents another example of this type of mapping, in which naive-physics understanding of force-dynamics situations underlies abstract notions of causation. Karmiloff-Smith (1992) suggests a developmental framework for this phenomenon called "representational redescription" which focuses on the gradual generalization of specific skills into abstract reasoning abilities. (Interestingly, it has been argued (Sheets-Johnstone 1990) that language itself slowly evolved from more innate and direct bodily expressions (e.g., crying for sadness).)

## 1.3 Approach—Converging Constraints

Ideally, of course, one would like to pin down precisely the algorithms and representations implemented by the portions of the nervous system underlying intentional activity. While such direct neurobiological grounding may be within reach in the domain of vision, such detailed understanding of low-level motor control has proved more elusive, and worse yet, hardly anything is known about the neurobiology of higher-level coordination of multi-step actions. Thus we must be satisfied with building *plausible* models which merely meet the constraints as we know them, and we must look beyond neurobiology for those constraints.

Foremost amongst these constraints is language. Language both reflects and influences our conceptual systems, and thus serves as a valuable window onto the processes we want to model. In particular, we will pay close attention to crosslinguistic variation in order to discover the range of concepts our model must be prepared to represent. However wide that range may be, however, we must confront the fact that the language-learning child is faced with such a bewildering abundance of perceptual and motor information that language learning would be impossible, if he did not know to ignore the vast majority of it. Similarly, the brevity of most utterances compared to the complexity of the situations they reference testifies to the amount of common ground assumed between speakers. Thus, it is crucial that we consult the psychological literature regarding natural *biases* and representational *abstractions* which characterize children's conceptual systems.

In sum then, we will attempt to integrate basic results from the following lines of research:

| |
|---|
| Encapsulation of motor functionality within motor synergies. |
| Coding schemes that appear to be used in the motor and premotor cortex. |
| Categorization of human grasp types. |
| General computational properties of neuronal networks. |
| Machine learning. |
| Crosslinguistic variations in verb meanings. |
| Perceptual attention biases and motor skills present at birth. |
| Patterns of child language development, including early errors. |
| Perceptual processing development and skill acquisition in early childhood. |

## 1.4 Contributions

### Semantics

Our task is formulated so that, we hope, it will naturally guide us towards a proper choice of features to represent, and a proper blend of hardwiring *vs.* learning. As a result, success at our learning task should allow us to make claims about the nature of the semantics of the class of verbs we consider. While we are not likely to make specific claims about neural structures, we do hope to justify the general notion of a multi-tiered representation which separates words, static features, finite-state schemas, and low-level continuous closed-loop control—a key feature of our model. Related semantics-learning projects include (Nenov & Dyer 1993; Nenov & Dyer 1994) and (Siskind 1992), though they do not share our grounding in motor control.

### Natural Language Understanding

We hope that understanding verb semantics at this level of fine detail will provide a strong foundation for understanding full sentences and stories. In particular we believe the active nature of our schema structure can support inference via *simulation*, a central aspect of Srini Narayanan's thesis work on understanding abstract metaphorical narratives based on schema semantics (Narayanan 1994). Demonstrating the *learnability* of verb$\Longleftrightarrow$schema mappings should mesh with Narayanan's schema-based story-understanding mechanisms to provide a fuller account of natural language understanding than either project alone.

### Knowledge Representation

By combining the constraint of learning linguistic mappings with the constraint of being able to drive execution, we hope to demonstrate that our representations of intentional activity—based on finite-state control augmented with feature extractors and constraint-enforcers—can bridge the procedural *vs.* declarative gap in Artificial Intelligence. Our system must achieve the behavioral advantages of SHRDLU's (Winograd 1973) procedures, while using a representation more amenable to recognizing, not to mention implementing, the subtle variations in verbs from the world's languages. Our current schema representation bears some resemblance to and is partially motivated by (Drescher 1991)'s schemas, which were designed to facilitate learning from experience.

## 1.5 Roadmap

The reader who is primarily interested in the cognitive bases for this work should now turn to Chapter 2, where we review relevant findings in linguistics, psychology, and neurobiology, and present results of our own pilot crosslinguistic data-gathering effort.

The reader who would rather skip directly to our proposed solution should turn to Chapter 3 where we overview our current model and work through a simple example. Chapters 4 through 6 then discuss aspects of the model in greater detail, including open design issues.

# Chapter 2

# Converging Constraints

Languages seem to vary widely, yet children learn their native tongue from a modest number of samples and quickly generalize their words correctly (or near correctly). Basic results in learning theory and AI show that such learning can take place only when considerable biases have already narrowed the scope of possibilities. What biases are built into humans? In this chapter we briefly review the cognitive science literature, in order to obtain a set of constraints that will guide us in solving our language learning task. We also present some preliminary results of our own linguistic data gathering effort.

Throughout the chapter we **boldface** important questions, some raised by our current computational model, to which we have not yet found the answers.

## 2.1 Crosslinguistics

### 2.1.1 Our own language survey

We are in the process of extracting intuitions, and eventually training data, from native speakers of various languages regarding hand-action verbs. The intent is to guide us in choosing features to incorporate into our model. The kinds of distinctions we seek are illustrated nicely by the following excerpt from Webster's:

> "TAKE, SEIZE, GRASP, CLUTCH, SNATCH, GRAB mean to get hold of by or as if by catching up with the hand. TAKE is a general term applicable to any manner of getting something into one's possession or control; SEIZE implies a sudden and forcible movement in getting hold of something tangible or an apprehending of something fleeting or elusive when intangible; GRASP stresses a laying hold so as to have firmly in possession; CLUTCH suggests avidity or anxiety in seizing or grasping and may imply less success in holding; SNATCH suggests more suddenness or quickness but less force than SEIZE; GRAB implies more roughness or rudeness than SNATCH."

We will try to translate such descriptions into settings of various parameters, something like:

| | **speed** | **force** | **security** | **precision** |
|---|---|---|---|---|
| **take** | any | any | any | any |
| **seize** | HIGH | HIGH | any | any |
| **snatch** | HIGH | LO/MED | any | any |
| **grasp** | any | any | HIGH | any |
| **grip** | any | any | HIGH | any |
| **grab** | any | any | any | LOW |
| **clutch** | any | HIGH | LOW | any |

Space prevents us from presenting a dictionary of all verbs we have collected, so instead we present just a few particularly intriguing non-English verbs.

- **PUDI (Tamil):** *Pudi* covers both obtaining an object, as well as continuing to hold an object. It connotes quickness in the first case, and exertion of force in the second case. It prefers the use of either a cupped palm supporting the object, or else a closed fist. Close English verbs are *catch*, *clutch*, and *restrain*.
- **VAIIE & PODU (Tamil):** Both verbs can refer to *put*ting an object down. Carefully-executed puts which ensure the object is placed securely use *vaiie*, although this verb is perhaps closer to English *keep*, since its prototypical case refers simply to maintaining an object in a given location, without expenditure of effort. Meanwhile *podu* connotes a careless put—indeed it includes throwing the object down. There does not seem to be an equivalent of *place*, connoting gentleness but focussing on the relocation of the object.
- **ILU & THALLU (Tamil):** These correspond roughly to pushing and pulling, however they connote a sudden action as opposed to continuous application of force and smooth movement. The only way to get this latter meaning is to suffix a directional specifier. Thus there is no way to indicate smooth pushing in an arbitrary direction.
- **HOL-DADAN & FESHAR-DADAN (Farsi):** These correspond to two different senses of *push*. *Hol-dadan* refers to moving an object away from oneself (it is actually closer to shove as it implies use of force; there is in fact no word for gentle or continuous pushing, other than the generic *move*-verb with a directional specifier). On the other hand, *feshar-dadan* refers to applying steady pressure to an unmoving object (e.g., pushing on the wall).
- **ZADAN (Farsi):** This word refers to a large number of object manipulations whose common character seems to be the use of quick motions. The prototypical *zadan* is a *hit*ting action, though it can also mean to *snatch* (*ghaap zadan*), or to *strum* a guitar (or play any musical instrument for that matter!).
- **MEET (Cantonese):** This verb covers both pinching and tearing. It seems to connote forceful manipulation by two fingers, yet it is also acceptable for tearing larger items where two full grasps are used.
- **LIC (Cantonese):** *Lic* serves as a generic *move* word when it is combined with various suffixes specifying direction. However it cannot be used alone, nor is there any other word to indicate general movement. Moreover, the verb *lic-jhu* means to keep still—not a move at all (*jhu* is not a negation).

- **DROP:** Neither Tamil nor Cantonese have a verb for gentle dropping. Both languages instead possess one verb for grip-release (i.e., *let-go*) which does not specify whether the released object is otherwise supported, and another verb for throwing down, which connotes use of force.

### 2.1.2 Crosslinguistic studies

Crosslinguistic variation in verbs is investigated most thoroughly in (Talmy 1985). A key result for *motion verbs*—a category that overlaps our domain—is that a given language will tend to encode only one of the following semantic categories (along with the fact of motion) in its root verbs: manner/cause, path, or figure. Other semantic categories can appear in verb roots as well, and Talmy has compiled a list which, considering only those items relevant to our simplified domain, gives us the following sources of linguistic distinction:

| |
|---|
| cause [European] |
| manner [Indo-European] |
| purpose (goal) [claims not used for motion verbs] |
| figure [Atsugewi] |
| path (in, out, up, down, past, through) [Polynesian, Semitic, Romance] |
| path+ground |
| polarity (occurrence *vs.* non-occurrence of key event) |
| phase (start, stop, initiate, finish) |
| aspect (one-way, full-cycle, iterative, continuous, gradient) |
| number of patients [Amerindian] |
| distribution of multiple patients (aggregate *vs.* linear) |
| direction (deictic) |

Properties importantly *not* encoded by verb roots include:

> resulting event subordinate to main event, ground alone, hedging, degree of realization, rate (Talmy claims it appears only in conjunction with other properties), spatial location of speaker or hearer, tense, speech-act (e.g. declarative *vs.* imperative)

An interesting property of action verbs also observed by Talmy is that goal can be rendered irrelevant in verbs expressing manner. For example, releasing and obtaining an object have distinct verbs (*put* and *take*) when no manner is prescribed, whereas *twist in* and *twist out* employ the same verb for two equally distinct goals.

Analogous studies are even better worked out for prepositions. For example, Eric Pederson has shown that within prepositions some spatial distinctions are more privileged than others, and has compiled this information into chart form for a dozen or so languages. Alas we do not know of similar work for early verbs in the hand domain. In a sense this project is a start at building such a chart.

Crosslinguistic variation has also been studied in the context of the “Whorf hypothesis” that cognitive skills can vary between speakers of different languages in a way that can be explained by the salience of certain concepts in one language but not the other. For example, it has been demonstrated that a subject's performance on puzzles conducive to thinking in absolute *vs.* relative coordinates, correlates with the extent to which the subject's language uses absolute *vs.* relative coordinate terms. More recently, Weist *et al.* (1994) have shown that Polish and Finnish children exhibit different relative skill on spatial *vs.* temporal reasoning tasks in a way which correlates with the relative "transparency" of spatial *vs.* temporal relation terms in the two languages. We do not know if there have been any similar results for verb-induced differences. In any case, these results are still controversial; see (Pinker 1994:ch. 3) for counter-arguments.

**Properties of some languages**

In Chinese (Mandarin), manner is especially important. For example, there is no general *hold.* Rather, one chooses a verb akin to *support* or *enclose*, depending on position of the hand.

Atsugewi is representative of a class of languages in which the type of the manipulated object is important. For example, Atsugewi's "move" verb depends on shape of object. An axial push (even if with an instrument) is distinct from a palmar push, which is also distinct from a push involving a quick hit. Hitting, in turn, is verbalized differently depending upon whether the hitter is swung or thrown at the object being hit. A very important question here is, what features do classifier languages use for verbs in our domain—in particular, do they tend to be just those features needed to successfully carry out the activity?

Japanese verbs draw a number of social distinctions involving either the actor(s), speaker, or listener (Talmy 1985). Such issues are beyond the scope of this thesis; we will simply use the appropriate verbs for parent-to-child speech. Japanese verbs may depend upon purpose to an unusual degree; we have yet to investigate this tip.

Mele encodes volitional *vs.* non-volitional status of a verb via a prefix.

Serial verb languages present some of the most interesting examples for our model. In such a language, the equivalent of *bring* might be expressed as a composite verb, e.g., *take-go-give.* Important questions to answer are (1) how many components are generally seen in a common verb; (2) how much variation of order is possible for a given set of components; and (3) how much does meaning depend on the component order?

(Levin 1993) offers an unusually comprehensive categorization of English verbs. Based on analysis of allowable argument structures (or "alternations"), the verb classes described in the work may provide useful clues to the underlying semantics of our verbs.

***** LINGUISTIC QUESTIONS *****

**QUESTION: Do action verbs preferentially refer to end-states, and especially attainment of goals?**

**QUESTION: Can action verbs' encoding of the coordination of "steps" of an activity, be captured by sequentiality, concurrency, branching, and looping (as opposed to, for example, relative duration, extent of overlap, etc.)?**

**QUESTION: Are the perceptual distinctions encoded by action verbs lim-**

**ited to those which could potentially affect the sequence of steps needed to carry out the action (and which are therefore contained in our schemas for executing those actions)?**

**QUESTION: Are there any demonstrated instances of action-verb semantics impacting motor skills, either developmentally or crosslinguistically?**

## 2.2 Psycholinguistic Development

### 2.2.1 First verbs

Clearly it is important for us to find out what children's first verbs are, both in general and for hand actions in particular. We have not yet found a direct source of this information, however, since the literature seems to catalog early nouns and function words, but discuss verbs only in the context of grammar acquisition (e.g., (Tomasello 1991; Slobin 1985)).

### 2.2.2 Innate perceptual biases and motor skills

Since Piaget's pioneering work on how conceptual systems could evolve from a nearly *tabula rasa* initial state, the thrust of much developmental psychology has been to show just how many cognitive skills are in fact present in children at birth (Mehler & Duphoux 1994). Object perception, object persistence, recognition of faces and emotions, and so on have been studied by Gibson and others. The role of inborn preferences in language learning has been considered by Spelke. Fisher (1994) has shown that when learning verbs valences, children are predisposed to assign subjecthood to *mobile* objects, even when motion and animacy are in conflict. Bates *et al.* (1991) consider how innate structures for language itself may have evolved, such as neonates' enhanced auditory resolving power for phonemes *vs.* other sounds.

Landau & Gleitman (1985), in a study of blind children's language development, offer some thoughts on what concepts might be "natural" for having a lexeme. Blind children learn the word *look* with little difficulty, and use their hands when given such a command. Sighted children, on the other hand, do not use their hands when blindfolded and asked to *look*. Landau hypothesizes that "look" is taken by all children to mean "explore with primary sense" and that this is a natural concept because of its obvious usefulness. Another claim is the naturalness of having just two, polar-opposite words for linear scales (e.g., *tall* and *short*).

Dunlea (1989) illustrates a host of differences between blind and sighted children's language that lead her to two conclusions. First, in blind children she finds a pattern of ascribing haptic-kinesthetic meaning to words which sighted children understand in a visual way, bolstering Landau's claim that an individual's primary means of perception holds a special place in language learning. Secondly, she finds that blind children tend to generalize much more narrowly. Tying this phenomenon to the blind child's generally limited knowledge of his surroundings, she makes the case that language learning depends crucially on worldly interaction.

### 2.2.3 Mechanisms underlying word and skill learning

Clearly, children's general cognitive abilities develop concurrently with their linguistic abilities. Various researchers have investigated the interactions between these developmental processes, resulting in the following ideas.

In the seminal work (Nelson 1973), word learning is claimed to be more a matter of concept *matching* than concept *formation*. Children first learn words about things they already care about and have concepts for. In this view, linguistic under- and over-generalization are products of the differences between the child's concepts and those of adults; as the child gains experience in the world his concepts take on adult form, allowing his language use to converge to adult form as well. Thus this work seems to deny language's role in guiding concept formation.

Huttenlocher *et al.* (1983) attack the problem by attempting to characterize the *types* of verbs learned at various stages of linguistic development. Their results are that (1) most verbs initially apply only to the child's own actions; (2) the first verbs that extend to cover others' actions are those involving overt motion; (3) later, verbs involving changed object state are properly extended; and (4) later yet, verbs requiring imputation of others' goals are mastered. This data clearly squares with the development pattern of increasing ability to map from the self to others, first physically and later mentally.

Gopnik (1981) essentially concurs with this view. This paper looks at functional words: "gone", "down", "there", "oh-dear", "no", "more", and argues that these increasingly complex words are learned as kids' physical strategies evolve. In "stage 1", the child acts randomly, but can make simple and arbitrary associations when reward occurs; at this point no functional words are learned. At stage 2, the child has learned the general idea that goals can be achieved by more than one action, and begins to apply all known actions to achieve a goal; words like "gone" and "down" which refer to goals rather than specific actions that achieve them are learned only at this stage. At stage 3, the child has begun to understand the effects of his actions, and thus can eliminate many possibilities during trial-and-error. Finally, by stage 4, the child has developed an effective world model, allowing planning of suitable actions through insight; now, the child can understand the rest of the functional words, which embody concepts of plan achievement, plan failure, plan repetition, etc..

Tomasello (1991) holds the view that language is built on general facilities and is interested in co-development for this reason. He suggests that change-of-state verbs emerge as object permanence sets in, and then grammar as an event-structure model of world develops. This work presents the "verb-island hypothesis", in which verb learning gradually leads to sentence understanding as the valences get figured out.

A potentially critical skill facilitating children's rapid vocaubular expansion is the identification of their language's semantic *patterns*. Choi & Bowerman (1991) demonstrate that 1–2 year old English- and Korean-speaking children have already generalized their language's (different) patterns regarding incorporation of path information into either prepositions or verb roots, respectively. Talmy's result (described in §2.1.2) that languages choose only one of manner/cause, path, and figure to encode in their motion verbs, provides another place where the child can simplify the verb learning task by grasping such patterns early.

Another indicator of how children ascribe semantics to new words is any patterns that can be found to their over- or under-generalizations. We have not yet looked into this literature.

In addition to determining the correct features, the language learner must also learn to *segment* time properly, i.e., determine the time window to which a verb refers. Gleitman has made progress in this area; we have yet to review this work.

***** DEVELOPMENTAL QUESTIONS *****

**QUESTION: What is the order of acquisition of the verbs pertaining to hand actions?**

**QUESTION: What are the common generalization patterns for verbs referring to intentional activity? Do children map these verbs to their own intentions right from the beginning? Is there a preference to associate verbs with goals rather than the parameters of actions, resulting in errors with manner verbs?**

**QUESTION: How do the child's increasing skills influence his verb usage?**

## 2.3 Motor Control

### 2.3.1 Behavioral observations

The muscle system has a forbidding number of degrees of freedom. We are looking for motor *synergies* which channel this freedom into a smaller number of useful motions. A step in this direction is the categorization of human grasps into a discrete set of grasp types. Many researchers have attempted this; a recent work is (MacKenzie & Iberall 1994) which further structures the grasp types using a small set of opposition primitives (pad, palm, side) in addition to the concept of "virtual fingers", in which the five fingers can be bound in various ways to two or three "roles" of grasping.

What environmental features determine grasp choice? Cutkosky & Howe (1990) have built an expert system that can predict appropriate grasp choices using a combination of object features such as shape and size, and task requirements such as power, precision, stability, or sensitivity.

While the planning of motor actions is not fully understood, we get a glimpse through experiments such as (Sternberg *et al.* 1978) in which the latency between being instructed to type a sequence of letters and initiation of the typing increased with the length of the sequence. Such experiments suggest that complex activity does in fact require the construction of an overall plan composing primitive actions. The composition, though, is not necessarily sequential. Oftentimes, primitive actions are performed concurrently. For example, (Arbib *et al.* 1987) describes finger preshaping that occurs while reaching for a mug. The preshaping varies depending on the mug size, and thus apparently is based on visual information.

### 2.3.2 Sub-cortical control

Clues to the parameterization of the motor system's synergies may also be found in the structure and function of the low-level nervous system which mediates muscle tension. Simple mechanisms such as the stretch reflex (Kandel *et al.* 1991:ch. 37) demonstrate the system's ability to exert local control of opposing muscles. A more complex synergy, the walking pattern, has been studied in the cat (Kandel *et al.* 1991:ch. 38). The intensity of a single signal from the brain-stem has been shown to modulate the speed of locomotion including transitions from walking to trotting to galloping. Such results support the notion that higher levels of motor control operate by adjusting small numbers of parameters on lower-level synergies.

A more general framework for synergies proposed by Latash (1993) is the *spring model*. Under this model, limbs are moved to their desired positions by providing each muscle involved with the final size and tension for the final configuration. The muscles then relax into this position according to the spring law. Such a system, however, faces a tradeoff between speed of motion and overshoot of the target position. (Jeannerod 1988) addresses this in his *triphasic model* for single joint control in which the limb is quickly moved toward the goal, is then pulled back, and then is finally eased forward again.

An important constraint on high-level motor control is the result that rapid motions be *ballistic*, i.e., they must execute without proprioceptive or sensory feedback. This is a simple consequence of the action potential time lag between limb and brain. Such open-loop actions require a more powerful representation in the top-level controller than would be required if all actions permitted continual feedback.

### 2.3.3 Motor cortex

Despite the surprising capabilities of the sub-cortical control system, cortical control is still required for tasks involving "voluntary" motion, memory, or novel concatenation of primitive actions. In the cortex, motor activity is mediated by the primary motor cortex, premotor cortex, and supplementary motor area. Primary motor cortex, located in the posterior portion of the frontal lobe adjacent to the somatosensory homunculus, whose structure it mimics, is responsible for both the initiation of movements and, in conjunction with the cerebellum, for their monitoring. The premotor cortex, located anterior to the motor cortex, is implicated in the mapping from a goal configuration into specific movements, and is believed to project these results to the motor cortex. Finally, the supplementary motor area, also located anterior to the motor cortex, is a particularly interesting region because it is active during *imagined* as well as actual movement. It's function appears to involve the coordination of multiple actions, e.g., maintaining balance while reaching. This structure has been inferred partly from deficit studies showing that motor-cortical lesions produce weakness, while premotor-cortical lesions impair purposeful movement and supplementary-motor-area lesions reduce coordination. See (Kandel *et al.* 1991:ch. 40) for a fuller description.

A major unresolved question is the coordinate system which motor cortex uses to plan movements. Kosslyn & Koenig (1992:ch. 7) review three alternatives that have been proposed. First, one might expect that "muscle coordinates" would be used directly. This hypothesis has been rejected in light of considerable evidence that many motor skills generalize across muscle systems; e.g., signing one's name on a chalkboard involves shoulder rather than wrist control, yet one's signature closely resembles the normal version. (This phenomenon, dubbed *motor equivalence*, was first observed by Bernstein.) A second possibility is that joint-angle coordinates are used. While computation of movements would be relatively simple in this space, the evidence does not favor its use. Most damaging is the observation that multi-joint movements tend to be smooth in body-segment coordinates, but convoluted in joint-angle coordinates. Thus, body-segment coordinates are apparently the most likely contender according to Kosslyn, who further claims that movements are planned by a two-step process in which the goal configuration and several *via points* are first identified, taking into account constraints such as obstacles, and then the actual motion involves smooth interpolation between these points.

What is represented at each point within the somatotopic map in motor cortex? Georgopoulos (1993), in a series of experiments, has determined that it is *direction* that is encoded. Furthermore he has found that a coarse-coding scheme is used; that is, the final limb direction is computed from the sum of the activity of a large number of broadly-tuned cells. Intriguingly, when told to move a limb at a certain offset from a visual stimulus, this population vector was found to *rotate* before movement began; furthermore, the rate of rotation matched that of Shepard's landmark work on visual mental rotation.

Supplementary motor cortex apparently is the locus of motor imagery. Jeannerod (1994) makes a case for motor imagery being similar to visual imagery, and argues further that a shared mechanism underlies both motor imagery and motor preparation. Such a result would enable the study of the motor control system via imaging tools such as PET, as the subject simply imagines performing an activity. The above paper also ties together a number of experiments and deficit studies into proposals on the motor system's representations of time, force/effort, goals, object affordances, and kinematic rules.

***** MOTOR-SYSTEM QUESTIONS *****

**QUESTION: What are the highest-level synergies known to control the behavior of the hand and arm?**

**QUESTION: Does the evidence support, at the cortical level of planned activities, multiple threads of control each of which activates synergies and tests perceptual feedback in a sequential manner?**

**QUESTION: How do goals index stored activity-sequences?**

**QUESTION: What fraction of the perceptual and proprioceptive information collected by receptors is used in local feedback loops, never making it to the cortex?**

## 2.4 Connectionist Modelling

"Connectionist modelling" (Feldman & Ballard 1982; Churchland & Sejnowski 1992) refers to a subfield of Artificial Intelligence whose mission is to abstract away the fundamental mechanisms of neural function which actually support cognition, and express them in computational terms, so that computer-science techniques can be employed to design and evaluate algorithms which are plausible models of human cognition. Some of the key principles this endeavor has identified are summed up below:

> PRINCIPLE: Neural networks are comprised of very many, fairly simple processors operating in parallel, exchanging simple messages, with no shared memory, slow updates, and a fixed topology in the short-term.

A variety of general techniques have been developed under this framework. A good overview is (Hertz *et al.* 1991). Algorithms exist for learning arbitrary mappings involving binary and numerical data, competitive learning, and various forms of classification. Recurrent networks have also been developed for recognition and generation of temporal patterns, although these are either computationally expensive, or limited in the kinds of patterns they can handle. In general,

> PRINCIPLE: Learning to categorize sequences is hard, especially when the sequences are long or of variable length.

Advantages of neural-network based algorithms include tolerance of noisy training data and straightforward parallelizabilty. However, they do not escape the following conclusion from both machine learning theory and practice (Mitchell 1980):

> PRINCIPLE: General learning techniques are usually inadequate for real-world tasks. Structure specific to the problem at hand must be built in in order to learn efficiently, or even to learn at all.

Within connectionist modelling, this structure needn't be precisely inspired by biology, but it should not violate any of the basic constraints of the first principle. Examples of highly structured connectionist work include (Goddard 1992) which presents a psychophysically accurate model of the recognition of moving light displays, based on "scenarios" of various gaits such as walking and running. These scenarios play a role similar to our schemas and may inform our design. As another example, Shastri & Ajjanagadde (1993) offer a connectionist account of logical inference matching psychological data on reflexive reasoning times and capitalizing on current thinking on the role of neural oscillations in representation (von der Malsburg & Singer 1988).

# Chapter 3

# Model Overview

## 3.1 Architecture

The architecture of our model is shown in Figure 3.1. Its structure has been chosen so as not to violate any of the constraints from Chapter 2, while facilitating computationally efficient verb learning. This bird's-eye view serves to illustrate our use of several distinct layers of representation, each designed to satisfy different task requirements.

The bottom layer of the model is the *schema layer*, which represents the skill base, i.e., knowledge of how to act to achieve various goals. The key constraint on schemas is that they must be *active*, i.e., capable of directly driving behavior. In the current design, schemas are enhanced finite state machines that encode the coordination of primitive actions including sequentiality, concurrency, branching, and propagation of parameters. Supporting this is a set of low level parameterized primitive actions roughly based on the data presented in §2.3. Our model will not be concerned with the implementation of these primitives, only how they are parameterized. This constitutes a claim that verb semantics will not depend on details at that level.

The top layer of the model is the *verb layer*. The representation here is rather straightforward and reflects our task's focus on isolated, atomic verbs. We simply include a "node" for each verb. (Though, it may prove important to change this representation to capitalize on compositionality of verbs, especially for languages such as Cantonese.) Note that we are not capturing any phonetics, vocal emphasis, or grammar.

Between these two is the *static feature layer*, significant since its presence is not dictated by the input/output of our task. This layer's job is to represent the salient characteristics of an execution of a given schema. Such characteristics might include the goal of the schema, the primitive actions performed, essential parameters such as speed, force, or direction, and a primitive form of aspect which can be derived from the flow of control through the schema, such as looping *vs.* linear progress. The purpose of this layer is to introduce very strong *biases* into the learning system, which are almost certainly necessary to support learning verbs from a small number of training examples. Of course, these biases are at the same time a potential liability, since they outright disallow the learning of certain classes of concepts over schema executions. In particular it will be inefficient to represent arbitrary details of the sequence of states visited by the schema.

The model includes several processes which transform information between these forms of rep-

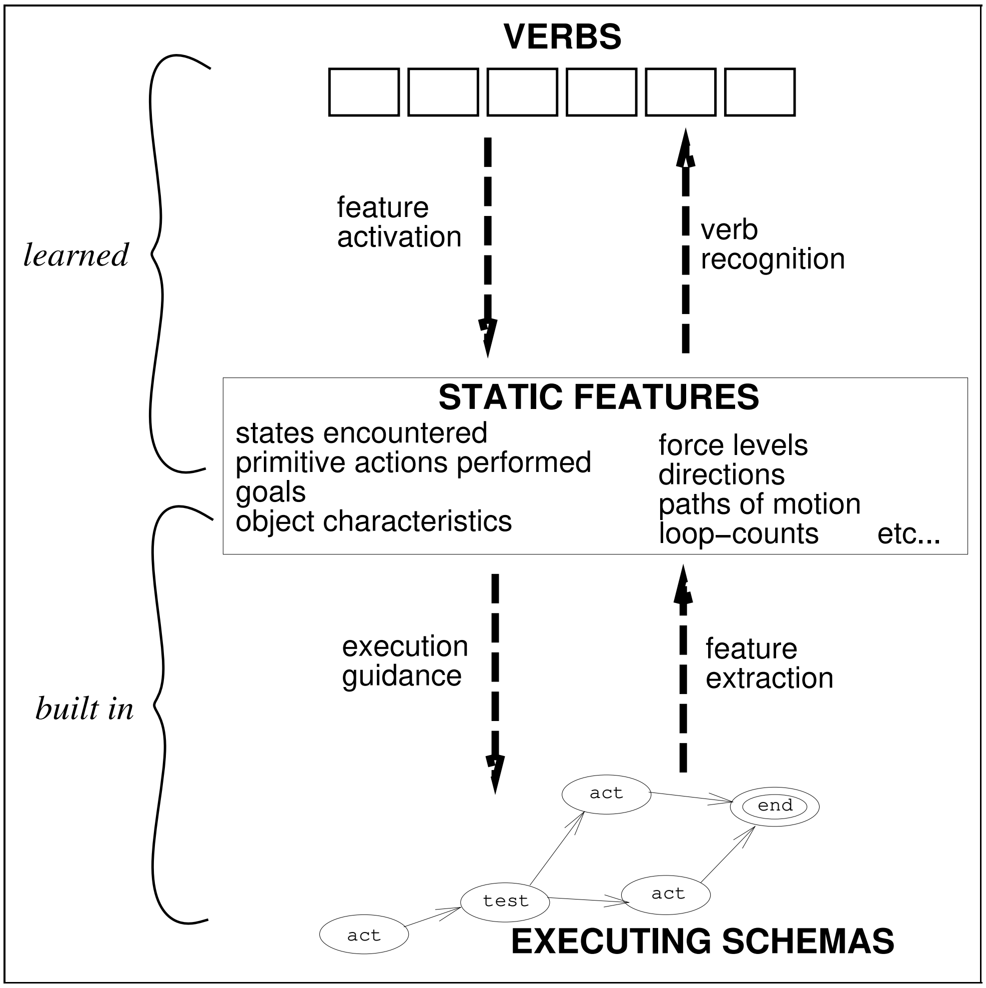


Figure 3.1: Top-level view of the model.

resentation. The *feature extraction* process is responsible for monitoring schemas as they execute in order to compute the summarizing features which must be represented in the static feature layer. Going in the other direction, the *execution guidance* process attempts to steer an executing schema in a direction compatible with the current contents of the static feature layer. In our model, both these processes are hardwired. The *verb recognition* process determines applicable verbs from the static feature layer, and this process is of course learned. The results of the learning must also, however, support the *feature activation* process.

## 3.2 Example

For methodological purposes, we will initially consider a simplified version of our task, in which we greatly limit the set of hand actions. This mini-task also serves as a concise illustration of how the complete model will work, and so we introduce it here. The mini-task domain contains just a tabletop and upon it a small box the size of a Rubik's Cube. For all training and test examples, the initial world state is the same—the hand is positioned at the box. The only actions considered are moves along the table surface. In English, the verbs that would apply are:

push, pull, shove, drag, yank, tug, slide, move

Figure 3.2 shows a bare-bones instantiation of our architecture to handle this mini-task. A key feature of the mini-task is that it does not require a large or complex schema set—every action can be considered an instantiation of a single `MoveAlongSurface` schema (shown at the bottom of the figure) which consists of choosing a means of contacting the box, followed by a single move primitive, with optional repetition. Thus, we are able to focus mainly on the parameterization of these two "synergies".

The table below lists a small set of features characterizing this space of actions. This set seems to capture the distinctions needed for the languages we have looked at to date. These features thus constitute the static feature layer for our mini-task. (A subset of these features can be seen in the center of Figure 3.2.) Most of the features are derived directly from the parameterization of the two primitive actions in the `MoveAlongSurface` schema; meanwhile the `aspect` feature is derived by monitoring the nondeterministic choice made from the `Repeat?` state.

| feature | values |
|---|---|
| contact-type | grasp, palmar |
| Aspect | iterated, continuous |
| Contact-at-end-state | kept, lost |
| XY direction | 0-360 |
| Horizontal force | positive reals |
| Vertical force | positive reals |

Because the initial world state is fixed and the `MoveAlongSurface` schema contains no branching based on perceptual tests, obeying verbal commands does not require a complex process of "guiding" schema execution through a variety of possible world configurations—rather, the verbal command completely determines the execution before it even begins. The parameter-based features simply instantiate the parameters of the schema execution, while the `aspect` feature determines

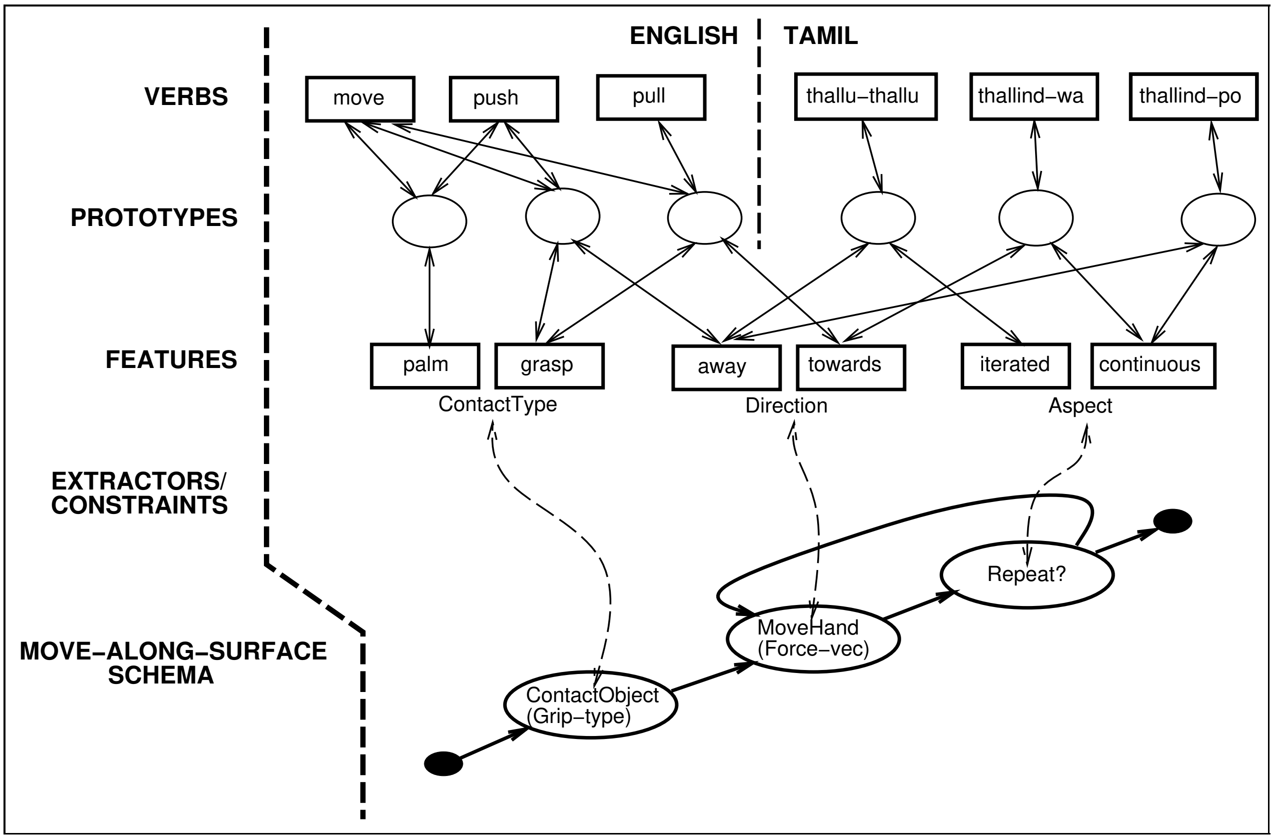


Figure 3.2: A simplified model handling moves along a tabletop. Layout follows Figure 3.1. (Presence of verbs from more than one language is for expositional purposes only; any instance of the model represents just one language.)

the behavior of the `Repeat?` state.

The bidirectional mapping from the feature layer to the verb layer can be learned by a clustering algorithm which produces a number of *prototypes*, designated by circles in Figure 3.2. Each prototype represents a conjunction of features which correlates with some verb, as well as the amount of tolerable deviation for numerical features. A verb may possess multiple prototypes. (In this simple mini-task, a verbal command causes a random prototype to be chosen.) Conversely, a prototype can activate multiple verbs. To illustrate, Figure 3.2 illustrates mappings for several English and Tamil verbs. The English *push* illustrates the use of multiple prototypes, while *move*, which includes pushing and pulling, shows how prototype can be shared. In Tamil, *thallu-thallu* refers to repetitive, jerky pushing; continuous pushing can only be indicated through the use of directional specifiers as in *thallind-wa* (towards self) and *thallind-po* (away from self).

There are a number of reasonable objections to this simplified example. Chapters 4 through 6 discuss the model in more detail, answering some (but not all) of these objections.

## 3.3 Issues

#### Hardwiring

It bears repeating that significant portions of the model are hardwired before learning begins—namely, the schema set, the static feature set, and the processes for mapping from one to the other. The motivation for this hardwiring, of course, is to shrink the space of possible verb concepts, so that learning can be made efficient. The rub, of course, is that the hardwiring might inadvertently render a verb impossible to learn by masking a crucial distinction. The answer to this problem is that, while the hardwired portions of the model are indeed fixed during a particular training episode, we are free to modify them when we see that the learning process is not working. Indeed, such modifications constitute the bulk of the scientific work ahead of us.

Eventually, if we want to model Whorf effects, we will have to permit modifications to at least the schema set. Despite the moving-target problem this will create, verb learning will still be aided by the informed pre-processing which the initial schema set and static feature set provide.

#### Static *vs.* dynamic

A natural objection is that our static feature vector is flawed precisely because it is static. After all, actions are dynamic and surely some of their temporal aspects are important in making linguistic distinctions. Our claim is that the relevant temporal information is in fact fairly restricted, so that we do not need the "shotgun approach" of presenting the verb-learning mechanism with full schema state-sequences for each training example. A better approach is to explicitly extract just the relevant temporal information, which should be brief enough that it can be presented to the learning mechanism statically. Such information includes properties of the world at the start and end states of the activity, and the number of times each primitive action executes. Additional temporal information is conveyed through parameters such as speed.

Nonetheless, it would be a reasonable experiment to try omitting all intermediate layers and attack the temporal learning problem of mapping directly from executing schemas to verbs. While this seems more difficult in general, it could in fact simplify learning in some cases, particularly for serial-order languages, for which the order of verb components may correspond to the order of primitive actions. In this case, we can leverage this temporal linguistic information by representing these components with separate nodes at the verb level which fire sequentially during schema execution.

### Pragmatics

Several pragmatics issues arise in our task. First, there is the potential for the same sequence of primitive actions and percepts to be labelled with different verbs, depending on which aspect of the activity is deemed most relevant by the labeler. However in our restricted domain it seems that the only source of such variability is differing assumptions about the actor's *goals*. But schemas very explicitly encode goals, so it is very natural for our model to learn to label a given sequence of primitive actions differently, depending on the goal which the sequence sub-serves. One difficulty remains, however, in that collecting training data via a somewhat crude animation package (see Appendix A) may make it difficult for labelers to infer correct goals.

A more serious pragmatic difficulty is that verbs semantics are modulated by modality, in this case the declarative *vs.* the imperative. When labelling training examples, subjects are likely to choose most-informative labels, for example labelling prototypical pushes and pulls with *push* and *pull*, reserving *move* for motions that aren't good pushes or pulls. Yet, when given a command to *move* an object, the model must be aware that it is acceptable to choose a prototypical push or pull when convenient. In other words, the system must learn that "pushes and pulls are moves". Such knowledge might be captured by a hierarchical version of the clustering algorithm we present in Chapter 6, though we have not worked out the details. Solving this problem may require modifications to the training set to include, in addition to the most-informative label for each activity, the set of acceptable labels.

### Image schemas

It is important to note that nowhere does the model include a complete description of the world state—the closest we come are the perceptual tests which schemas perform in order to achieve their goals. This represents a claim that the workings of the motor-control system are the central component in the semantics of our verbs. An alternative approach would be to combine this motor-control information with other kinds of perceptual information into a *image-schematic* representation which would lie between the schema layer and static feature layer in our model. Then, the static features would be extracted from these image schemas. With such a design we might be able to extract salient features which we cannot capture by inspecting schemas alone. It would also provide a clearer path for integration with Regier's model. On the downside, however, extraction of these features would probably be more difficult. An important open question is whether such an approach is necessary for the verbs we are considering.

# Chapter 4

# Representation

Recall that our model has three levels of representation: a schema-like representation for controlling intentional activities; a static-feature-vector for representing salient semantic features; a grandmother-cell-like representation for verbs themselves. This chapter describe these representations, their rationale, and open issues.

## 4.1 Schemas

Essentially, the schema structure encodes control flow, including sequence, concurrency, testing and branching, looping, and random choice. Such a structure depends on being able to find a level of description of motor and sensory functionality that can be squeezed into this finite-state structure. In this section we describe the schema structure in detail and then outline an initial selection of primitive motor actions and sensory discriminations with which to construct our working schema set. See Figure 4.1 for an example schema.

### 4.1.1 Finite state machines

In the simplest case, a schema is simply a finite state machine. One such machine exists for each goal. The schema has a start state and final state, and between them it may have:

- **action-states**, labelled with a primitive action, and having a single outgoing transition; and
- **test-states**, labelled with a test involving a primitive percept, and having two or more outgoing transitions, each labelled with a possible test result.

Note that in such a schema, the path of execution from start to finish is determined entirely by the external world, which supplies the results in test-states.

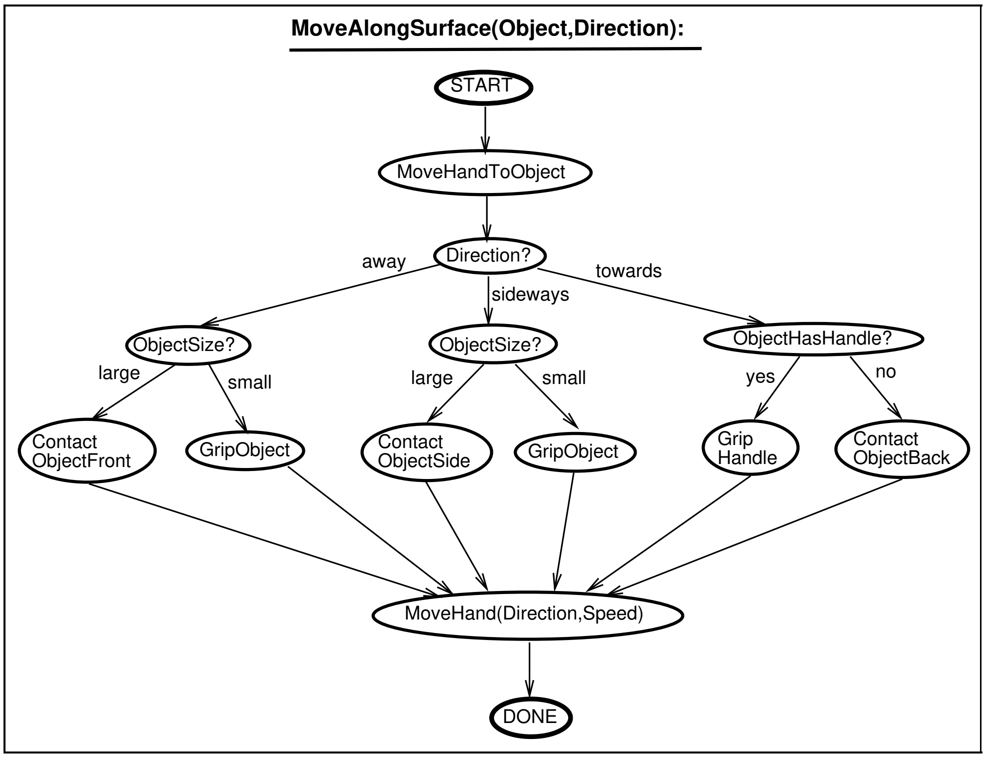


Figure 4.1: A sample schema.

### 4.1.2 Enhancements

Building from this starting point, we add constructs to represent nondeterministic branches, concurrent threads, hierarchical invocation of sub-schemas, and parameterization.

#### Nondeterminism

Many goals may be attainable in a variety of ways even from the same initial world state. To model this, we introduce

- **or-states**, with no label, and having multiple unlabelled outgoing transitions.

During schema execution, a random branch is taken, unless the model is obeying a verbal command and has learned that only some subset of the choices is applicable to the verb, in which case the random choice is made from this subset.

An extension of this would be to have probabilities on the transitions to indicate "suitability". Then, when under verbal command, these probabilities would be modified in some way, as opposed to restricting the choices discretely.

### Concurrency

For realistic behavior (and perhaps to capture certain linguistic distinctions), it is probably necessary to model concurrently executing primitives—e.g., the opening of the fist while the hand moves towards a mug (Arbib *et al.* 1987). To model this we add:

- **fork-states**, with no label, and having multiple unlabelled outgoing transitions, with the effect that each transition gives rise to a concurrent thread of control; and
- **join-states**, with no label, and having a single unlabelled outgoing transition, which do not pass control along their outgoing transition until *all* of their incoming transitions have passed control to them.

No guarantees are made regarding synchronization of actions or tests along concurrent threads.

### Hierarchy

We can eliminate much redundant schema structure by encoding frequently-used functionality into a "subschema" which may be called directly by an action-state of another schema. The caller state does not pass along control until the callee completes. This is the same abstraction seen in Augmented Transition Networks (ATNs) commonly used in parsing.

So long as the schema set is fixed and we are only learning language, extending schemas to be hierarchical is probably nothing more than an easily implemented convenience. However, if we *do* get into schema modification or learning, extraction of common subsequences to generate or modify the hierarchy would become central.

An important consequence of hierarchical schemas is that they imply a hierarchical structure of goals and subgoals; now schemas start to look like plans. This permits learning verbs which refer to the execution of certain schemas, but only as subgoals of a larger task. An example of such a verb is Tamil *nahathu* which can refer to moving objects in various ways, but only when done to clear the object out the way as part of some other task.

### Parameterization

Recall that our primitive actions and tests are parameterized. We take advantage of this to parameterize schemas as well. In other words, a schema invocation can include top-level parameters, which are applied to the primitive actions during execution. At issue is what kinds of *transformations* are allowed when applying parameters in this way. A simple version would be for each internal parameter to have a default value, plus an optional pointer to a top-level parameter with a "gain" indicating the degree to which the top-level parameter should modulate the internal parameter. For example, this would permit a schema's top-level speed parameter to vary the speed of *some* of its steps while other steps execute quickly in all cases.

Parameterization of this sort can be implemented by augmenting each schema with enough globally accessible, read-only memory to hold the top-level parameters. However, even more powerful forms of parameterization are possible, such as allowing states to modify this global memory, or to pass information to the next state. Such a facility would allow primitive actions to be parameterized by the results of earlier perceptual tests.

With this enhancement in particular, we are moving well beyond the realm of finite state machines, and there is a price to pay. The space of possible schemas grows astronomically with parameterization, tremendously complicating the task of learning schemas from experimentation. The problem can only be attacked with assumptions about the kinds of transformations that are important.

### 4.1.3 Sensorimotor primitives

Our set of sensorimotor primitives should meet several criteria. First of all, they are intended to model motor synergies in the human motor system. While the detailed structure of these synergies is not known, the primitives should reflect the data reviewed in §2.3. An automatic sanity check on the plausibility of our primitives is that we must be able to map them to simulator commands (see Appendix A about our use of an animated human body simulator for data collection and evaluation). Secondly, our primitives should be as flexible, i.e. parameterized, as possible, so a smallish set of them can support a large variety of activities when composed into schemas.

Our perceptual-tests may utilize perceptual features including contact, direction-to-object, object size, force, and degree-of-control.

Motor primitives will include grasps of several types (tripod, cylindrical, palmar contact), ballistic and controlled hand movements, compliant motion, and static application of force. Parameters of these primitives include force, speed, degree-of-control, direction, and size.

### 4.1.4 Alternative: Petri Nets

Another potential model of motor control is condition-action rules. We can think of such a system as having a bunch of copies of each primitive action, each of which is *guarded* by a number of conditions. Conditions may be features of the world, or triggers from other parts of the control system. Petri Nets provide a formalism for such a representation. Advantages include a natural way of handling concurrency, and the potential to handle unusual conditions without having to clutter the representation for the common cases. Extensions to handle hierarchical nets and parameterized actions should be similar to the finite state machine case. One disadvantage of Petri nets is that, since they are strictly more powerful than finite state machines, learning them will be more difficult. However, it would be possible to work with a restricted class of Petri nets to facilitate learning, and this is arguably cleaner than extending finite state machines.

## 4.2 Static Features

The static-feature layer summarizes the execution of a schema into a relatively small number of linguistically salient features. Recall that its purpose is to simplify learning by turning a fundamentally temporal learning problem into a static one. Structurally, the static feature layer is trivial: it is simply a vector of nodes each representing a single feature. These nodes are grouped into *segments* of related features, and various coding schemas are employed.

We now list the most important feature-segments and some sample values for each. We also list some reasonable encoding schemes for each segment, such as 1-of-N, many-of-N (a.k.a. distributed), real-valued, vector-valued, and coarse-coded.

- **Goals:**

| | |
|---|---|
| `obtain` | `move-3d` |
| `forfeit` | `restrain` |
| `move-along-surface` | `deform` |

  These are represented by a bit-vector segment, with a single bit set to 1 for any given schema execution.

- **Object information:**

| | |
|---|---|
| `size` | `weight` |
| `direction` | `distance` |

  Fairly straightforward real values, except for direction, which could be vector-valued, or nominal-valued (e.g., up, down, towards, away). This choice has not been made.

- **Parameters:**

| | |
|---|---|
| `precision` | `stability` |
| `speed` | `force` |

  All real-valued.

- **Primitive actions:**

  (see §4.1.3)

  Here we again have a bit vector, with one bit assigned to each primitive action; i.e., parameters are ignored. Bits will be set for all primitive actions that are performed during schema execution.

- **Aspect:**

| | |
|---|---|
| `once` | `one-cycle` |
| `iterated` | `continuous` |

  As these are mutually exclusive, they will employ 1-of-N encoding just as for goals. Of course, this is a very limited subset of aspectuals.

- **World state:**

| | |
|---|---|
| `support` | `contact` |
| `possession/control` | |

  It may prove essential to partially represent the sequentiality of schema execution. A limited way of doing this would be to include just the initial and final states of the world; this list is a suggestion for what features might be important at these two points in time.

### 4.2.1 Open Issues

We must consider how to indicate features which are "not applicable" to a given schema execution (and which are, consequently, "don't cares" during command execution).

An interesting idea is to introduce a hierarchy of features. For example, we might have goals such as `move` along with `move-along-surface` and `move-vertically`. Such a "multi-resolution" structure would provide a richer substrate for verb learning and allow varying degrees of verbal control over schema execution.

We could choose to represent real-valued features by coarse-coding, in which a small number of nodes, each standing for a fixed value, represent an input value by signalling the closeness of their fixed value to the input value. (The L, M and S cones in the retina are a good example of this kind of coding.) This pre-clustering of real values would be useful in the case where we have a reason to expect that certain values will be important linguistically.

A major challenge for the static feature layer emerges from the case of verbs with multiple prototypes. A mechanism is needed by which the system can choose the best prototype based on the initial configuration of the world. Better yet would be a mechanism that allows switching to a new prototype when the schema corresponding to the best prototype fails during execution.

## 4.3 Verbs

Lastly, we have a very simple representation for verbs in which a so-called grandmother cell represents each non-compound verb. Note that this precludes making any use of phonetic similarity, and also vocal emphasis unless we treat different emphases of a single verb as distinct verbs.

For compound verbs, we could continue with this simple strategy. But if we instead have nodes for each verb component and represent compound verbs by concurrent firing of multiple component nodes, then we can take advantage of shared structure and speed up learning. (But see §3.3 for a discussion of potential advantages of *sequential* activation of the component nodes.)

Functionally, each verb node or component node has incoming connections summarizing the evidence for or against the verb's applicability in the current context. Verb nodes also possess competitive links with each other, so that the system can choose the single most appropriate verb. In the compositional-verb case, it would be useful if the system could learn to group the component nodes (e.g., into "roots", "directional specifiers", etc.). Then, component nodes would compete only within their component group.

# Chapter 5

# From Schemas to Features, and Back

The representational divide discussed in Chapter 4 requires that we build two translation mechanisms—one to translate executing schemas into static features to facilitate labelling, and another to to translate static features into guided schema execution, for execution of verbal commands. This chapter describes these two processes, which are hardwired into the model and tailored to specific characteristics of the schema and static-feature layers.

## 5.1 Feature Extraction

Each node in the feature vector is controlled by a feature extractor. Here we give the flavor of this process with a few examples.

The extractors for goals are the simplest. Each goal feature is controlled by an OR mechanism which receives an excitatory connection from the final-states of all schemas that achieve the goal. (See Figure 5.1.) Goal extraction will become more complicated if we have hierarchical schemas, or parameterized goals.

Extractors for aspect are slightly more complex. We can distinguish between `once` and `iterative` aspect by treating specially those transitions which go connect a "deep" state to a more "shallow state", where deep and shallow refer to a depth-ordering of the schema from the start-state. When control passes through such a transition, we have detected a loop and so the `iterative` feature can be turned on. Conversely, if the `once` feature is on by default, passage through such a transition should turn it off. (See Figure 5.2.) This simple mechanism will not make all needed aspectual distinctions. In particular, it would be impossible to distinguish a single cycle through the loop, from the loop being repeated many times. A solution is to build a feature extractor which sums repeated activations of a transition and then thresholds it.

It is not clear exactly what form the integrative-parameter feature extractors should take. Structurally, they must be connected to all action-states which specify the parameter. Mathematically, they might compute a simple average, or a weighted average if we can find a basis for the weighting.

Extracting the set of primitive actions performed is a trivial matter involving an excitatory connection from every action-state to the appropriate primitive-action feature bit.

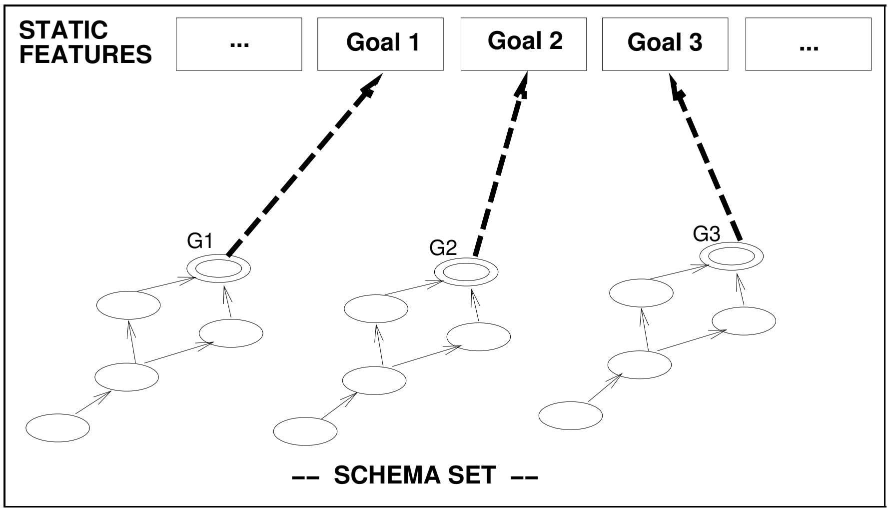


Figure 5.1: Goal extractor. Each goal feature receives an excitatory connection from the final-states of all schemas that achieve the goal.

---

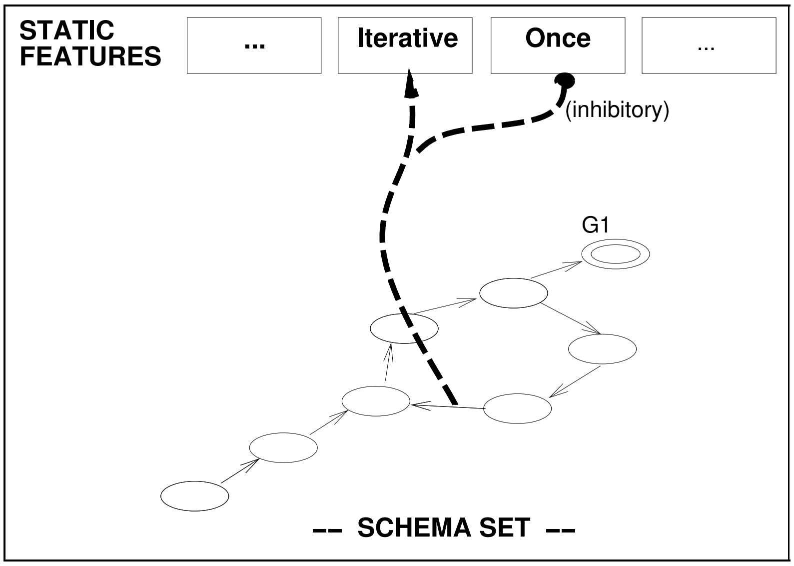


Figure 5.2: Aspect extractors. Transitions to states with a lower depth-ordering have an excitatory link to the `iterative` feature, and an inhibitory link to the otherwise-active `once` feature.

---

We have not yet considered how the world-state features will be set.

An interesting question is whether the feature-vector should be regarded as being set only after a schema has completed execution, or whether it should gradually be "filled in" *during* execution. In this latter case, we would be able to address issues of exactly when during performance of activity the model is able to produce the verb—more data to match, another constraint on design.

## 5.2 Guided Schema Execution

Goal features must index the start-states of all schemas which achieve the given goal.

Aspect features, e.g., the `iterative` feature, must provide priming to or-states. Finding the appropriate or-states to prime involves a backward search from the transitions that activate the `iterative` feature.

Integrative features such as `overall-speed` must connect to all action-states with speed-parameterized actions. A question is, how to represent the extent to which a given action should be influenced by this feature. (For example, there may be some steps in a schema which should always be executed slowly, even when overall the activity is being done quickly.)

It is not clear if there is a big role for primitive-action features in guiding execution. Potentially, if a verb strongly selects for a certain primitive action to occur, this information might be used to influence or-node branches.

**Issue**

In the midst of all this custom-tailored machinery, there is at least one identifiable function which has general applicability. Often it will be the case that the the static feature vector wants to "prime" a given state in a schema, that is, encourage the schema to visit that state during execution. This comes up, for example, in both the aspect and primitive-action features. The only means of coaxing a schema towards a given state is to manipulate the decisions made at or-states "upstream" in the schema execution. Thus, it would be useful to have mechanism to determine the appropriate or-states for a given primed state. This could be done via spreading activation.

### 5.2.1 Saying "I can't!"

Sometimes the world will be in a state which does not permit the carrying out of a verbal command without violating some portion of the semantics of the verb. We will need a mechanism for detecting this case and reporting it. This issue has not yet been addressed.

# Chapter 6

# Verb Learning

This chapter considers the task of learning the mappings between static features and verbs. Currently, we are working on a solution based on clustering, which we just sketch here. Issues involving compositional verbs are also discussed, as well as some avenues by which learning performance may be improved.

## 6.1 A Clustering Algorithm

One might consider an off-the-shelf algorithm such as backpropagation for our task, except for two hitches. First, children are able to use a new verb in a reasonable and possibly even correct manner after just a few hearings. While we will not at first require our system to mimic this learning pattern, we prefer algorithms that appear to be extendible to handle this condition, ruling out algorithms based solely on incremental weight changes. Second, as we have noted before, our learned mappings must be bidirectional in order to support verbal-command execution.

For simplicity, we'll only require our mappings to produce the *best* output for a given input, rather than, for instance, reporting the full probability distribution over outputs. That is, a given static-feature vector containing motor-features $\hat{m}$ and world-state $\hat{w}$ must produce the single verb $v$ satisfying

$$\underset{v}{\operatorname{argmax}} \{P(v|\hat{m}, \hat{w})\}$$

while for command execution, a given verb $v$ and initial world-state $\hat{w}$ must produce the motor features $\hat{m}$ satisfying

$$\underset{\hat{m}}{\operatorname{argmax}} \{P(\hat{m}|v, \hat{w})\}\,.$$

With discretization of the numerical features, it would be painful but possible to do the learning by directly storing frequencies in a matrix indexed by feature-vectors and verbs. However this provides no opportunity for generalization, and does not scale well.

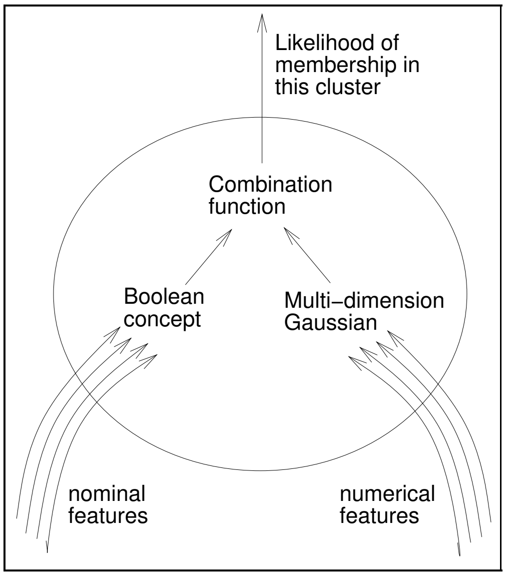


Figure 6.1: Structure of a prototype unit.

So instead we will investigate clustering of the static feature vectors. Clustering involves deriving a set of *prototypes*; new data is classified by finding the closest-matching prototype. In our system, these prototypes are intended to correspond to the linguistic notion of word-sense. So, for example, a verb may have multiple prototypes; *push* might have the two prototypes:

1. hand moves object away from self, by whatever means
2. hand contacts object with palm, force is exerted continually, but object does not move.

The class of clustering algorithms includes competitive learning; EM and its efficient approximation K-means; Autoclass; Learning Vector Quantization (LVQ); and Adaptive Resonance Theory (ART). Clustering algorithms can be distinguished along two axes: (1) whether the number of prototypes is fixed or variable; and (2) whether the data is labelled with output classes or not. In our task the number of prototypes is not known in advance, requiring our algorithm to determine an appropriate number. However our data is labelled, so will be performing *partially supervised* clustering; two examples with different verbs surely belong to different prototypes, but examples sharing a verb may or may not belong to the same prototype.

Figure 6.1 depicts the "prototype units" which our algorithm will construct. Their structure is made complex by the presence of "nominal" (i.e., unordered, such as the set of goals) as well as numerical features in the static-feature vector. (We will refer to these as the "nominal subvector" and "numerical subvector" for shorthand.) The units handle this complication by having two "sites"—a multi-dimensional Gaussian site for the numeric features, and a Boolean concept site for the nominal features.

We first present an *offline* algorithm based on the abstract notion of model-merging as presented in (Stolcke & Omohundro 1994); then we consider the *online* case.

We begin learning by creating a prototype unit for each training example. In each, the Boolean concept is set to the conjunction of the features in the nominal subvector, and the Gaussian is very highly peaked around the values in the numerical subvector. Then, the algorithm repeatedly considers merging pairs of prototype units, stopping when no "good" candidate merges exist. The suitability of a candidate merge is a function of (1) the overlap of the Gaussians of the two prototypes; (2) the overlap of the Boolean concepts; and (3) the relative importance of these overlaps. A number of evaluation functions are possible.

If a merge is made, the new prototype unit's Boolean concept is set to the disjunction of the two nominal subvectors, and the Gaussians are summed (weighed by the number of examples each prototype represents) and normalized. The new prototype unit also records the number of training examples it explains, namely the sum of the counts in the merged prototype units.

Once trained, recognition is performed by applying the test feature-vector in parallel to all prototype units. Each evaluates the degree of match using the same function used for evaluating candidate merges (i.e., the test vector acts like a prototype unit for a single training example). The verb nodes receive this evidence from the prototype units and the most activated verb node constitutes the system's output.

For command execution, the system first selects those prototype units corresponding to the given verb. From these, the maximum likelihood prototype is determined, where the likelihood is a function of three quantities:

1. the number of training examples assigned to the prototype unit ("commonness")
2. the peakedness of the prototype unit's Gaussians and the number of satisfying assignments of its Boolean concept ("selectivity")
3. the similarity between the prototype unit's world-state features and the current world state ("applicability")

Once a winner is chosen, the motor-control features of the static feature vector are set as follows: the nominal features are each set to a random satisfying assignment of the winner's Boolean concept, and the numerical features are each set to the mean of the winner's Gaussian. During schema execution, these features will guide the choice of parameters and may affect branching.

We would like to work out an incremental, online version of the algorithm. Besides opening up the possibility of modelling the *trajectory* of children's verb-learning, an incremental algorithm would solve the scalability problem of starting with one prototype unit per training example. The online version will work by attempting to incorporate new training examples into the current set of prototypes, constructing new prototype units only when this fails. A key issue in such an algorithm is how to determine this threshold of incorporability so as to yield a model of good size. (For reference, this is the "vigilance" parameter in ART.)

The above algorithm is not capable of sharing prototype units between verbs. Sharing is desirable since it reduces the number of prototype units needed, and can reduce the number of training examples needed for learning. For example, if after learning *push* and *pull*, the system were amenable to reusing these verbs' prototype units, it could learn *move* very quickly. However, this extension requires a criterion for deciding to adopt an existing prototype in this way.

## 6.2 Leveraging feature-grouping

The segmented structure of the static-feature vectors can be leveraged to introduce biases that will speed learning. A given language is likely to exhibit a regular pattern of encoding certain types of features more often than others. Learning would be aided, therefore, by including a mechanism which attempts to learn this preference ordering over the feature-vector segments for the language being learned. Once correctly learned, the system will have a helpful bias for learning new verbs. This mechanism has yet to be designed.

## 6.3 Handling compositional verbs

Detecting preference ordering will be even more useful if we represent compositional verbs with component nodes instead of verb nodes. But first the system must learn *component grouping* (e.g., roots, *vs.* directional specifiers *vs.* object specifiers) by noting which components never appear together in a single verb. A simple benefit of establishing this grouping is that it permits eliminating competition between component nodes in different groups. More interestingly, it permits us to learn a preference ordering of feature-vector segments *within each component group.* In other words, the system could potentially learn that components in "group 1" tend to encode directional features, while "group 2" components tend to encode goal, etc.

## 6.4 Query Generation

Our model permits access to the schemas from which it must generate verb semantics, and so it is in a good position to be able to identify holes in its current mappings, due to lack of appropriate training examples. Thus we should look into algorithms to generate appropriate queries to fill such gaps. (Pragmatically, this of course requires us to have a subject present during training.)

Similarly, we might get considerable benefit from a small number of *negative* training examples. The source for these would be the model's own failures when tested for either recognition or execution.

# Appendix A

# Methodological Details

This appendix deals with several methodological aspects of the project, including our means of legitimately collecting training data and validating our results through the use of an animated body simulation system.

## A.1 Data Collection and Evaluation Criteria

Collecting sufficient numbers of legitimate <`verb, schema-execution`> pairs for the relevant verbs from a reasonable number of languages poses some difficult problems. It is unnatural, and hence error-prone, to ask subjects to label schema executions directly, since this would require an analytical decision as opposed to the quick, intuitive, decisive kind of labelling that we really want. It is also unacceptable for the experimenter to translate a subject's demonstration of an activity into a schema execution, since the process is neither verifiable nor repeatable.

We have a related problem when it comes to evaluating the results of the learning process. Evaluating the system's recognition abilities can be done by holding out a portion of the collected training examples and seeing how the system labels them. But evaluating the system's command-execution abilities is more difficult, as subjects must assess the appropriateness of the resulting schema execution.

Both problems can be solved, however, by use of a graphic simulation package. If a simulator has a sufficiently accurate human body model that mimics the functionality of our proposed primitive actions and perceptual tests, then we can collect data by driving the simulator with a collection of sample schema executions and having subjects label the visual presentation of the activity being carried out. Similarly, for command-execution, the simulator can provide the perceptual test results needed to steer the resultant schema through its execution in the current world state, and and also provides a graphic output for evaluation by the subject.

## A.2 *Jack*

We believe we have found such a simulator in *Jack*, a human body animator developed at University of Pennsylvania (Badler *et al.* 1993). Though aimed at the design factors engineering market, its sophisticated body models and high-level commands should render the interfacing between simulator and schema traces manageable. Visually, the simulator presents shaded polygonal images (the standard human has several hundred polygons, updating at a few frames per second). Hopefully this spatial and temporal resolution will be adequate.

The body model, including the hand, is fully jointed. Joints have adjustable ranges of motion as well as force limits. Coordinated movements are considered, and the system performs the inverse kinematics and dynamics necessary to support realistic motion. For example, if the simulator is asked to move the hand forward, the arm will follow, eventually the shoulder will turn and the torso will lean, and finally the body will actually step forward. Strength-guided motion is built in, as is collision avoidance. We are not yet certain, however, if all these features work simultaneously, especially in (quasi-)real-time. We have not yet had a chance to evaluate *Jack*'s set of movement commands and how easily our primitive actions can map to them. An important question is how, and whether, the simulator can be given overlapping movement commands.

Currently we know less about what perceptual information is available outside the simulator. For high-level percepts, the simulator does allow access to a database of objects in the scene, including their locations and properties. If we can also access velocities of all segments and forces at all joints, then we will be able to compute most of the percepts we expect to need in our schema set.

*Jack* limits the potential realism of the objects in our toy world. We should be able to handle the following. The body may be seated at a table and so the tabletop provides a base. Gravity exists. Object properties include shape (simple constructive solid geometry), weight, and surface characteristics such as coefficient of friction. Objects may also possess a few salient parts such as handles.

Interestingly, the designers of *Jack* have already built a natural language interface embodying some of our ideas. They also have a schema-like representation for complex tasks (Esakov & Badler 1990) which we will be looking into.

# Bibliography

ARBIB, MICHAEL A., THEA IBERALL, & DAMIAN LYONS. 1987. Schemas that integrate vision and touch for hand control. In *Vision, Brain and Cooperative Computation*, ed. by Michael A. Arbib & Allen R. Hanson. Cambridge, MA: MIT Press.

BADLER, NORMAN I., CARY B. PHILLIPS, & BONNIE LYNN WEBBER. 1993. *Simulating Humans*. New York: Oxford University Press.

BAILEY, DAVID, 1992. Inference through imagery: Understanding simple sentence sequences in $L_0$. Unpublished paper from class in natural language understanding at UC Berkeley.

BATES, ELIZABETH, DONNA THAL, & VIRGINIA MARCHMAN. 1991. Symbols and syntax: A Darwinian approach to language development. In *Biological and Behavioral Determinants of Language Development*, ed. by Norman A. Krasnegor, Duane M. Rumbaugh, Richard L. Schiefelbusch, & Michal Studdert-Kennedy, chapter 2. Hillsdale, NJ: Lawrence Erlbaum Associates.

BERLIN, BRENT, & PAUL KAY. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press.

CHOI, SOONJA, & MELISSA BOWERMAN. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. In *Lexical and Conceptual Semantics*, ed. by Beth Levin & Steven Pinker, 83–122. Amsterdam: Elsevier Science Publishers.

CHURCHLAND, PATRICIA A., & TERRENCE J. SEJNOWSKI. 1992. *The Computational Brain*. Cambridge, MA: MIT Press.

CUTKOSKY, MARK R., & ROBERT D. HOWE. 1990. Human grasp choice and robotic grasp analysis. In *Dextrous Robot Hands*, ed. by Subramanian T. Venkataraman & Thea Iberall, chapter 1. Springer-Verlag.

DRESCHER, GARY L. 1991. *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. Cambridge, MA: MIT Press.

DUNLEA, ANNE. 1989. *Vision and the Emergence of Meaning: Blind and Sighted Children's Early Language*. Cambridge University Press.

ESAKOV, JEFFREY, & NORMAN I. BADLER. 1990. An architecture for high-level human task animation control. In *Knowledge-Based Simulation: Methodology and Application*, ed. by Paul A. Fishwick & Richard B. Modjeski, chapter 9. New York: Springer-Verlag.

FELDMAN, JEROME A., & DANA BALLARD. 1982. Connectionist models and their properties. *Cognitive Science* 6.205–254.

——, GEORGE LAKOFF, DAVID R. BAILEY, SRINI NARAYANAN, TERRY REGIER, & ANDREAS STOLCKE. 1994. $L_0$—the first five years of an automated language acquisition project. *AI Review* 8. Special issue on Integration of Natural Language and Vision Processing, to appear.

——, GEORGE LAKOFF, ANDREAS STOLCKE, & SUSAN WEBER. 1990. Miniature language acquisition: A touchstone for cognitive science. Technical Report TR-90-009, International Computer Science Institute, Berkeley, CA. also in the Proceedings of the 12th Annual Conference of the Cognitive Science Society, pp. 686–693.

FISHER, CYNTHIA. 1994. Who's the subject? The meaning of structure in preschoolers' interpretations of novel verbs. In *Proceedings of the Child Language Research Forum*. University of Chicago Press.

GEORGOPOULOS, APOSTOLOS P. 1993. Cortical representation of intended movements. In *Neuroscience: From Neural Networks to Artificial Intelligence*, ed. by Pablo Rudomin, Michael A. Arbib, Francisco Cervantes-Pérez, & Ranulfo Romo. Springer-Verlag.

GODDARD, NIGEL, 1992. *The Perception of Articulated Motion: Recognizing Moving Light Displays*. University of Rochester dissertation.

GOPNIK, ALISON. 1981. Words and plans: Early language and the development of intelligent action. *Child Language* 9.303–318.

HERTZ, JOHN A., ANDERS KROGH, & RICHARD G. PALMER. 1991. *Introduction to the Theory of Neural Computation*. Redwood City: Addison-Wesley.

HUTTENLOCHER, JANELLEN, PATRICIA SMILEY, & ROSALIND CHARNEY. 1983. Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review* 90.72–93.

JEANNEROD, MARC. 1988. *The Neural and Behavioural Organization of Goal-Directed Movements*. Oxford University Press.

——. 1994. The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences* 17.

JOHNSON, MARK. 1987. *The Body in the Mind*. The University of Chicago Press.

KANDEL, ERIC R., JAMES H. SCHWARTZ, & THOMAS M. JESSELL (eds.) 1991. *Principles of Neural Science*. New York: Elsevier, third edition.

KARMILOFF-SMITH, ANNETTE. 1992. *Beyond Modularity*. Cambridge, MA: MIT Press.

KOSSLYN, STEVEN MICHAEL, & OLIVIER KOENIG. 1992. *Wet Mind*. New York: Free Press.

LAKOFF, GEORGE. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

LANDAU, BARBARA, & LILA R. GLEITMAN. 1985. *Language and Experience: Evidence from the Blind Child*. Harvard University Press.

LATASH, MARK L. 1993. *Control of Human Movement*. Human Kinetics Publishers.

LEVIN, BETH. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

MACKENZIE, CHRISTINE L., & THEA IBERALL. 1994. *The Grasping Hand*. North-Holland. Advances in psychology, vol. 104.

MEHLER, JACQUES, & EMMANUEL DUPHOUX. 1994. *What Infants Know: The New Cognitive Science of Early Development*. Cambridge, MA: Blackwell. Translated by Patsy Southgate.

MITCHELL, THOMAS M. 1980. The need for biases in learning generalizations. Technical Report CBM-TR-117, Dept. of Computer Science, Rutgers University.

NARAYANAN, SRINI, 1994. A computational model of online narrative interpretation. Ph.D. Thesis Proposal, forthcoming.

NELSON, KATHERINE. 1973. *Structure and strategy in learning to talk*. University of Chicago Press. Monograph of the Society for Research in Child Devleopment 38(1-2).

NENOV, VALERIY I., & MICHAEL G. DYER. 1993. Perceptually grounded language learning: Part 1 — a neural network architecture for robust sequence association. *Connection Science* 5.

——, & ——. 1994. Perceptually grounded language learning: Part 2 — DETE: A neural/procedural model. *Connection Science* 6.

PINKER, STEVEN. 1994. *The Language Instinct*. New York: William Morrow and Company.

REGIER, TERRY, 1992. *The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization*. Computer Science Division, EECS Department, University of California at Berkeley dissertation. Available as Technical Report TR-92-062, International Computer Science Institute, Berkeley.

ROSCH, ELEANOR, CAROLYN MERVIS, WAYNE GRAY, & DAVID JOHNSON ADN PENNY BOYES-BRAEM. 1976. Basic objects in natural categories. *Cognitive Psychology* 8.382–439.

SHASTRI, LOKENDRA, & VENKAT AJJANAGADDE. 1993. From simple association to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences* 16.

SHEETS-JOHNSTONE, MAXINE. 1990. *The Roots of Thinking*. Philadelphia: Temple University Press.

SHEPARD, ROGER N. 1992. The perceptual organization of colors: An adaptation to regularities of the terrestial world? In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. by Jerome H. Barkow, Leda Cosmides, & John Tooby, chapter 13. Oxford University Press.

SISKIND, JEFFREY, 1992. *Naive Physics, Event Perception, Lexical Semantics and Language Acquisition*. Massachusetts Institute of Technology dissertation.

SLOBIN, DAN ISAAC. 1985. *The Crosslinguistic Study of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

STERNBERG, S., S. MONSELL, R. L. KNOLL, & C. E. WRIGHT. 1978. The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In *Information Processing in Motor Control and Learning*, ed. by G. E. Stelmach, 117–152. New York: Academic.

STOLCKE, ANDREAS, 1994. *Bayesian Learning of Probabilistic Language Models*. University of California, Berkeley dissertation.

——, & STEPHEN OMOHUNDRO. 1994. Best-first model merging for hidden Markov model induction. Technical Report TR-94-003, International Computer Science Institute, Berkeley, CA.

TALMY, LEONARD. 1985. Lexicalization patterns: semantic structure in lexical forms. In *Language typology and syntactic description: grammatical categories and the lexicon*, ed. by Timothy Shopen, chapter 2. Cambridge University Press.

——. 1988. Force dynamics in language and thought. *Cognitive Science* 12.49–100.

THRUN, SEBASTIAN B. 1992. The role of exploration in learning control. In *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, ed. by D. White & D. Sofge. Van Nostrand Reinhold.

TOMASELLO, MICHAEL. 1991. *First Verbs*. Cambridge University Press.

VON DER MALSBURG, C., & W. SINGER. 1988. Principles of cortical network organization. In *Neurobiology of Neocortex*, ed. by P. Rakic & W. Singer. London: John Wiley & Sons Ltd.

WEIST, RICHARD M., PAULA LYYTINEN, & JOLANTA WYSOCKA, 1994. Whorfian reflections in development. Poster presented at the 1994 Stanford Child Language Research Forum.

WINOGRAD, TERRY. 1973. A procedural model of language understanding. In *Computer Models of Thought and Language*. New York: W. H. Freeman.