C. Project Summary

Computer systems that could understand natural language would enormously broaden the range of possible computer users and applications. Although there have been some successful specialized natural language understanding (NLU) systems, these have all been limited in scope, brittle, and very expensive to build and maintain. The advances of the last decade in statistical language processing have been very useful for some tasks, but can not on their own yield understanding. NLU requires extensive semantic knowledge and sophisticated inference methods.

In previous NSF work, we have constructed a pilot system capable of understanding news stories in the domains of politics and international economics. The stories were taken from newspaper articles in these domains. When presented with a pre-parsed version of these narratives, the system was able to generate commonsense inferences consistent with the input. The pilot system used a novel model of actions and events that combines an executable action representation with belief propagation and metaphor mappings (Narayanan 1999b; Narayanan 1999a). The representation integrated an extended Stochastic Petri Net (Ciardo *et al.* 1994) representation (x-schemas) designed to model distributed dynamic systems with complex coordination, concurrency and resource constraints with Temporal Bayes Networks (TBN) (Murphy 2002) for inference.

The news understanding pilot and subsequent extensions (Chang *et al.* 2002) exhibited an architecture capable of capturing many of the rich linguistic and conceptual subtleties needed for real NLU. There are obviously significant scientific and engineering challenges involved in extending the architecture to large scale tasks.

One of these challenges, extending the inference module, is the main focus of the current proposal. A formalism integrating x-schemas for action and TBN for inference appears to be an attractive methodology for the construction and analysis of intelligent systems. For large scale systems, we will need to utilize more modular and structured probabilistic networks, usually called Probabilistic Relational Models or PRM (Pfeffer 2000; Pfeffer & Koller 2000; Getoor *et al.* 2001). Our current efforts at the computational level are directed towards a complete integration of PRM and x-schemas, which we call Coordinated Probabilistic Relational Models or CPRM. These can be seen as extending an ongoing community effort to add more sophisticated temporal and control capabilities to probabilistic inference models (Pearl 1988; Boyen & Koller 1998; Anderson *et al.* 2002; Jordan 1999; Lerner *et al.* 2002).

A PRM is a qualitative dependency structure and a set of parameters quantifying the conditional dependency relations. These are much more scalable than flat TBN because the dependency model quantifying the various domain relations is given at the frame level. CPRM add the control flexibility of Petri nets extend the formalism to cover a wide range of action models. A central research question is how to produce efficient realizations of CPRM and our initial ideas on this are part of the current proposal. In addition to developing the theory of CPRM, we will implement and make generally available, a software package that efficiently implements the theory.

This is part of a larger effort on scalable NLU and we will also work on integrating the CPRM inference engine with the other two critical components: a parser and interfaces to large external knowledge sources. This integration effort articulates with other past NSF work on the FrameNet linguistic knowledge project and on Embodied Construction Grammar.

We propose to develop the theory of CPRM, implement a generally available package that efficiently realizes the theory, and demonstrate its inferential capabilities on a scalable version of the existing pilot understanding system. As parallel efforts to build semantic analyzers and on wide-coverage ontologies and knowledge compendia mature, we will be ready to exploit these resources for large-scale applications in NLU. The combination of these advances, along with those from other labs, should permit the development and testing of scalable NLU systems within a few years.

D. Project Description

1 Introduction

Computer systems that could understand natural language would enormously broaden the range of possible applications and users. Although there have been some successful specialized Natural Language Understanding (NLU) systems, these have all been limited in scope, brittle, and very expensive to build and maintain. The advances of the last decade in statistical techniques for Natural Language Processing (NLP) have been very useful for some tasks, but can not on their own yield understanding. NLU requires extensive semantic knowledge and sophisticated inference methods.

We propose to extend existing knowledge representation and probabilistic reasoning methods and demonstrate how this can support scalable NLU systems. In particular, we will develop the theory and implementation of Coordinated Probabilistic Relation Models (CPRM). These combine the control flexibility of Petri nets (Murata 1989; Ciardo *et al.* 1994) with the expressive and inferential power of the Bayesian extension of relational data bases (Pfeffer 2000; Pfeffer & Koller 2000). This is part of a larger effort on scalable NLU and we will also work on integrating the inference engine with the other two critical components: a parser and an interface to large external knowledge sources.

The proposed work arises as a large scale extension of a previous pilot system for understanding news stories involving international economics. Understanding such stories requires sophisticated reasoning about actions and events. The world's languages have a variety of grammatical and lexical devices to *construe*, *focus*, *direct attention* and *control inferences* about actions and events (Vendler 1967; Moens & Steedman 1988; Langacker 1991; Verkuyl 1993; Steedman 1996; Narayanan 1997b). Consider the meaning of *stumbling* in the following newspaper headline "Indian Government stumbling in implementing Liberalization Plan". Clearly, the speaker intends to specify that the liberalization plan is experiencing some difficulty. Moreover, the grammatical form *is + VP-ing* suggests that the difficulty facing the plan is *ongoing* and the final outcome of the plan is indeterminate. Compare this to the subtle meaning differences with grammatical and lexical modifiers on the same root verb such as *has stumbled* or *starting to stumble*. Most readers are likely to infer after reading this sentence that the government's liberalization policy is likely to fail, but this is only a default causal inference that is made in the absence of information to the contrary. Finally, how does *stumble*, whose basic meaning is related to spatial motion and obstacles get interpreted in a narrative about international economic policies?

Several decades of research in cognitive science have established that there are powerful primary schemas underlying much of language and thought (Langacker 1987; Talmy 1988; Talmy 1999; Lakoff & Johnson 1980; Lakoff 1987; Johnson 1987; Slobin 1997). The NTL project (http://www.icsi.berkeley.edu/NTL) has been developing computational realizations of this theory for both scientific and applied purposes, including a program that interprets simple causal narratives in the domains of Politics and Economics. The stories were taken from newspaper articles in these domains. When presented with a pre-parsed version of these narratives as input, the system was able to generate common sense inferences consistent with the input. The pilot system used a novel model of actions and events that combines an executable action representation with belief propagation and metaphor mappings. The system was able to interpret narratives such as the one below about India's march toward liberalized economics from the New York Times.

> In 1991, in response to World Bank pressure, India boldly set out on a path of liberalization. The government loosened its strangle-hold on business, and removed obstacles to international trade. While great strides were made in the first few years, the Government is currently stumbling in its efforts to implement the liberalization plan.

The system was able to draw correct inferences from a pre-parsed version of this story, using a variety of techniques that we will extend in the proposed project. For concreteness, we will first describe the relevant features of the pilot system.[1] Section 3 will present our general idea for scaling up such systems and Section 4 contains the technical core of the proposal, implementing CPRM. Section 5 compares CPRM with logical theories of action. Section 6 shows how our existing system, and hence CPRM, can be used to support understanding of arbitrary semantic frames[2] Finally,

---

[1] The Pilot System, KARMA (Section 2) describes previous work done under NSF grant IRI-9619293 with Feldman as PI Narayanan as a supported CS graduate student at UC Berkeley.

[2] FrameNet related work described in Section 6.1 is funded under the NSF FrameNet project NSF ITR/HCI 0086132, where Narayanan is a Co-PI.

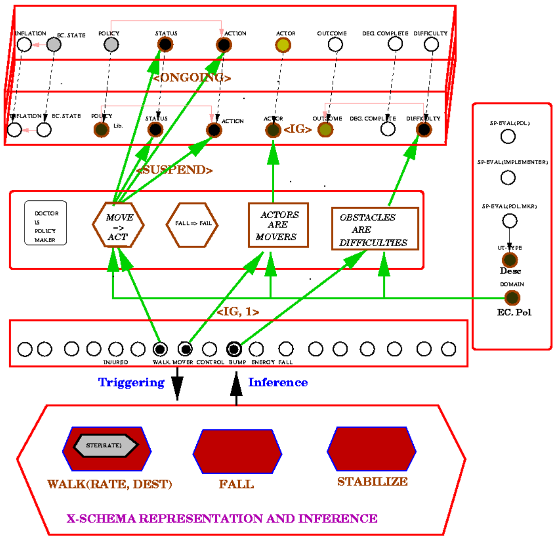


Figure 1: KARMA System Architecture. **Metaphor maps** project values from **source domain** action and event simulations (see Figure 2) to **target domain** evidence on one or more slice of the target domain temporally extended Bayes net (TBN) (see Figure 3).

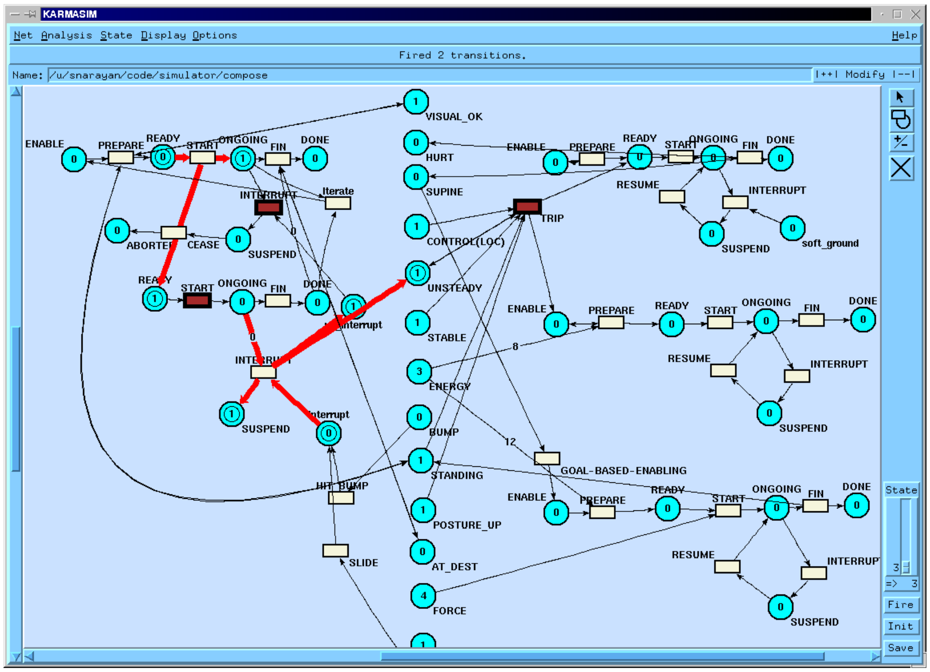
in Section 6.2, we describe how the proposed inference module can articulate with efforts on scalable parsers and knowledge compendia, under development at ICSI and elsewhere.

2 KARMA: A Pilot Inference System for Understanding

In the example above, note that institutions are conceptualized as causal agents, causes as forces, actions as motions, and goals as states in a spatial terrain. These mappings are part of a cross-linguistic metaphor system called the Event Structure Metaphor (Lakoff 1994) which is the general name for projections from the concrete experiential domain of forces and spatial motion (source domain) to the abstract domain of causes, actions, and events (target domain). Following from the fact that institutions are conceptualized as agents, specific causal events are attributed as effected by or affecting the institution; such as apply pressure, respond to pressure, loosen strangle-hold, remove obstacles, stride, and stumble. Commonsense inferences that are required for interpreting the article often *rely* on our experience of force dynamics and motion in space. For instance, the inference that stumbling *leads to* falling can felicitously be transferred to the abstract domain of economic policy through a conventionalized metaphor that *falling* ↦ *failure*. This enables the interpreter to conclude that the government is likely to fail in its liberalization plan. Many other inferences rely on the source domain (consider the implications of strangle-hold).

Figure 1 shows the architecture of the implemented pilot inference system, KARMA (Narayanan 1999a). The specific hypothesis motivating the design is that the meaning of motion and manipulation terms is grounded in patterns generated by our sensory and motor systems as we interact in the world. Systematic metaphors project these features onto abstract domains such as Economics enabling language to use motion terms to describe abstract actions and processes. The implemented system has three main components, namely the **source domain**, the **target domain** and the **metaphor maps**. The source and target domains are based on a model of action that is able to meet the representational requirements and support the kinds of inferential processes inherent in language understanding and is the focus of this proposal.

The central idea behind the model is that the reader interpreting a phrase that corresponds to a motion term is in fact performing a mental simulation of the entailed event in the current context. The basic idea is simple. We assume that people can execute schemas with respect to structures that are not linked to the body, the here and the now. In this case, actions are not carried out directly, but instead trigger simulations of what they would do in the imagined situation. The physical world is modeled by other schemas that have i/o links to the schema representing the planned



Fired 2 transitions.

Name: /u/snarayan/code/simulator/compose

Figure 2: The source domain is a rich action and event simulation environment. Shown is a fragment of the network simulating the situation of an agent's *ongoing* WALK being interrupted by an *obstacle* resulting in a situation where the agent is *about to* TRIP and *maybe* FALL.

action.

In the pilot implementation, source domain structure is encoded as connected motion schemas (see Figure 2). The model of the source domain is a dynamic system based on inter-schema *activation*, *inhibition* and *interruption*. In the simulation framework, whenever an executing x-schema makes a control transition, it potentially modifies state, leading to asynchronous and parallel triggering or inhibition of other x-schemas. The notion of state as a graph marking is inherently distributed over the network, so the working memory of an x-schema-based inference system is distributed over the entire set of x-schemas (see Figure 2). This control and simulation regime remains central to the proposed CPRM design. Of course, it is also intended to model the massively parallel computation of the brain (Feldman 1990).

An important and novel aspect of our source domain representation is that the same system is able to respond to either direct sensory-motor input **or** other ways of setting the agent state (such as linguistic input). This allows for the same mechanism to perform simulative reasoning and generate inferences from linguistic input as well as be used for high-level control and reactive planning. There is now robust biological evidence to support the view (Gallese *et al.* 1996; Grafton *et al.* 1996; Tanji 1994) that planning, recognition and imagination share a common representational substrate. We believe that this is an important aspect of embodiment allowing the same mechanisms to reason as well as react. The structure of the pilot abstract domain (the domain of international economic policies) encodes knowledge about Economic Policies. The representation must be capable of a) representing background knowledge (such as US is a market economy), b) modeling inherent target domain structure and constraints (high-growth may result in higher inflation), and c) be capable of computing the impact of new observations which may from direct input ("US economy is experiencing high-growth"), or from metaphoric (or other) inferences ("Economy stumbling"). Furthermore, these different sources of evidence have different degrees of believability, and the representation must provide a framework for their combination. For all these reasons, we chose to represent the target domain as a Bayes Net.

The pilot model of the target domain entails multiple copies (up to 4) of a temporally extended Bayes net (Dean & Wellman 1991), representing different time slices. The structure of the target domain for three temporal slices of the Bayes network is shown in Figure 2. Within a single temporal slice, the nodes of the network correspond to economic variables which can take on different values. For instance, in we have a node in the target network (Figure 2) corresponding the the economic actor which can be instantiated to be the US government, IMF, Indian

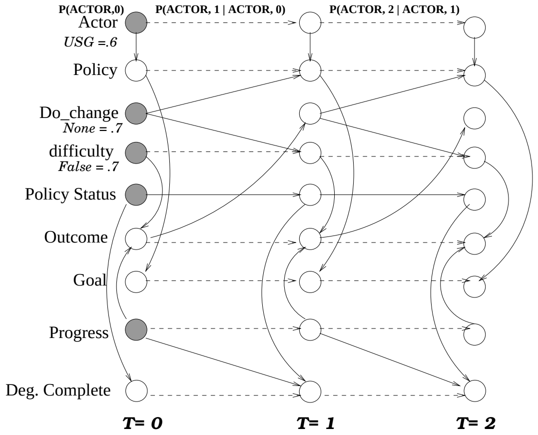


Figure 3: The target domain is a temporally extended Bayes net (TBN).

Government, etc. Links within a single time slice model the probabilistic dependence between variables. For instance, there is a link between the actor variable and the policy variable, which models the fact that if we knew the actor in question (US Government) we would have a good idea of the policy (free-market economy). The strength of this belief is quantified as the conditional probability table $P(Policy|Actor)$. Links between nodes at different time slices encode the conditional probability of a variable's value at time t, given its value at $t-1$. For instance, the link $P((Actor,1)|(Actor,0))$ (ref. to the top of Figure 2) results in the conditional probability table (CPT) that corresponds to the probability of a specific actor being instantiated at time $t = 1$, given the value of the actor at time $t = 0$. From such local conditional probability tables, BELIEF PROPAGATION algorithms (Jensen 1996; Jordan 1999) compute the global posterior probabilities for the entire network propagating influences backward and forward in time. As we will see, a major barrier to scaling the pilot system is the inefficiency of very large unstructured bayes networks.

In the pilot model, metaphor maps connect the x-schema based representations to the Bayes net representing knowledge about international economics (see Figure 1). Such maps *project* specific results of x-schema executions by projecting specific source domain values to the target domain by asserting new *evidence* at one or more time slices of the temporally extended Bayes net.

A story represents the specification of a *partial* trajectory over epistemic states. This is simulated by clamping some of the Bayes network nodes to specific values. The remaining features are estimated using known target domain inter-feature correlations as well as metaphoric projections from the embodied general knowledge (x-schemas). Metaphoric projections of x-schema executions may clamp target features to specific values (by placing new *evidence* on the target domain Bayes net.

Comprehending a story corresponds to finding the set of trajectories that best satisfy the constraints of the story and are consistent with the domain knowledge. This may involve *filling in* missing values or placing new evidence on the Bayes net. The resultant target network state becomes a prior for processing the next input at stage $t = 2$. Background knowledge is encoded as the network state at $t = 0$. Target inferences can go forward and backward in time in the estimation of the most probable explanation of the input story.

In the pilot system, the embodied domain theory had about 100 linked x-schemas, while the abstract domain theory is a relatively sparse net of about 40 multi-valued variables with at most 4 temporal stages. It also encoded about 50 metaphor maps from the domains of *health* and *spatial motion*. These were developed using a database of 50 $2-3$ phrase fragments from newspaper stories all of which have been successfully interpreted by the program. Among the inferences made were those related to goals (their accomplishment, modification, subsumption, concordance, or thwarting), resources, aspect, frame-based inferences, perspectival inferences, and inferences about communicative intent. (Narayanan 1997a; Narayanan 1999a; Chang *et al.* 1998; Bailey *et al.* 1998; Chang *et al.* 2002) report on the different types of inferences produced by the system.

In summary, our results suggest that a large proportion of commonplace descriptions of abstract events seem to project embodied, familiar concepts onto more abstract domains such as economics and politics. This allows non-experts to comprehend and reason about such abstract policies and actions in terms of more familiar and universal embodied concepts. The fact that the metaphoric inferences are context-sensitive, immediate, and defeasible set up fairly strong representational requirements for a metaphor interpretation system. The structured probabilistic representation coupled with the rich action semantics of x-schema based simulation enables our model to capture subtle contingency relations between events necessary for routine commonsense inference.

3 Scalable NLU systems

Inference for NLP applications such as event tracking or question answering requires a rich computational model of the temporal and causal structure of events. For instance, knowing the *stage* of the process (interrupted walk due to a stumble event), gives valuable *predictive* information (may fall) as well as *presuppositional* information (a walk has started and was ongoing). Of course, this information is *probabilistic* (the agent is only **likely** to fall), *abductive* (Hobbs *et al.* 1988) (the agent was *probably careless*), *dynamic* (is about to trip), *context sensitive* (the type and severity of stumble in combination with the agent's energy and ability to stabilize will determine whether the agent falls), often *metaphoric* (as in economies stumbling, sliding, crawling, sprinting), and may incorporate the specific *perspectives* of the various participants.

The news understanding pilot and subsequent extensions (Chang *et al.* 2002) exhibited an architecture capable of capturing many of the rich linguistic and conceptual subtleties needed for real NLU. There are obviously significant scientific and engineering challenges involved in extending the architecture to large scale tasks. Some critical challenges include building a 1) building a scalable semantic analyzer to extract semantic relations from text, 2) the availability of and integration with large scale ontologies and wide-coverage language resources and 3) a computational framework that can support the rich set of inferential requirements in a scalable manner.

One of these challenges, extending the inference module, is the main focus of the current proposal. A formalism integrating extended Petri nets for action and temporal Bayes networks for inference appears to be an attractive methodology for the construction and analysis of intelligent systems. For large scale systems, we will need to utilize more modular and structured networks, usually called Probabilistic Relational Models or PRM (Pfeffer 2000; Pfeffer & Koller 2000; Getoor *et al.* 2001). Our current efforts at the computational level are directed towards a complete integration of PRM and x-schemas, which we call Coordinated Probabilistic Relational Models or CPRM. These can be seen as extending an ongoing community effort to add more sophisticated temporal and control capabilities to probabilistic inference models (Pearl 1988; Boyen & Koller 1998; Anderson *et al.* 2002; Jordan 1999; Lerner *et al.* 2002).

Section 4 details our technical approach to solve this problem. Section 5 relates our approach to the other formal theories of actions and events. Section 6 outlines how the proposed work articulates with ongoing parallel efforts (by us and others) to address a) the issue of large-scale knowledge compendia and b) the issue of open-domain semantic extraction (other critical barriers to large-scale NLU).

4 A Scalable and Expressive Model of Actions

Reasoning about structured stochastic dynamic systems requires modeling coordinated temporal processes and complex, structured states. A significant amount of work has gone into different aspects of overall problem.[3] Figure 4 maps out the space of relevant probabilistic modeling and inference techniques along three basic dimensions (extended from the description in (Anderson *et al.* 2002)). The dimension along the x-axis (left-right) depicts the increasing expressiveness of the action model. The y-axis (vertical going up) corresponds to increasing the complexity of the state representation. The z-axis (into the plane) corresponds to increasing the richness of the overall representation. The origin of the space is an unstructured probabilistic state vector representation with no explicit temporal or relational information.

Moving to the right along the x-axis, we get to linear temporal models of sequences. Markov Models (MM) are the most widely used technique to model such simple sequential processes. They have achieved considerable success in a variety of domains (speech, computational biology). However, Markov models (including Hidden Markov Models (HMM) which are properly subsumed under DBN (or TBN) (Murphy 2002)) are fairly inflexible and representationally

---

[3] In all these cases, we can have continuous variables as well as discrete ones. For the purposes of this exposition, all the comments here apply to both types of states and actions.

inadequate as models of actions.[4] Specifically these representations are unable to model and reason about central aspects of actions such as *concurrency*, *synchronization* and *resources*. Moving further right, we arrive at a set of well developed graphical modeling approaches designed to model distributed dynamic systems with complex coordination, concurrency and resource constraints. These representations are based on Stochastic Petri Nets (SPN) (Ciardo *et al.* 1994), are used widely in modeling in many domains (such as networks, distributed systems, computational biology).

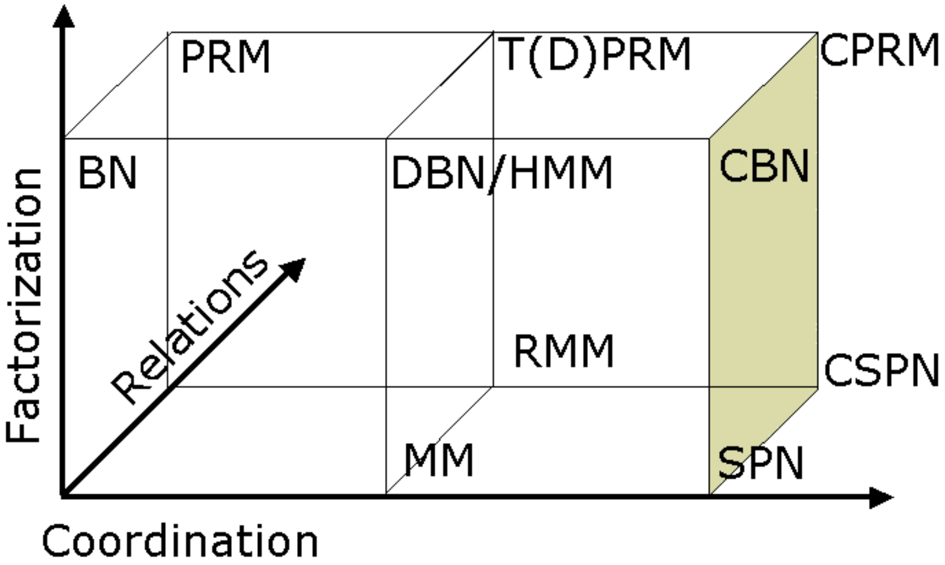


Figure 4: Probabilistic Models and Inference Space

One of the main drawbacks of MM and SPN based representations is the inability to represent states with complex internal structure. Moving from the origin up along the y-axis, we have Factor Models, Markov Random Fields (MRF) (the undirected version) and Bayes Nets (BN) (the directed version) all of which make conditional independence assumptions to factorize the joint probability distribution of the state variables into a compact product form. Temporally extended Bayes Nets (TBNs, also called DBNs) model each time step in a sequence as a BN and links between state variables at different time steps capture the temporal dependencies between variables. TBNs are thus able to model simple sequences and structured state variables. Moving rightward, CBNs (Narayanan 1999b) combine the expressive action modeling framework provided by the SPN (or CSPN (SPN with types as in x-schemas (see section 4.2)) based representation with the ability to model complex states provided by the BN framework. This model of action and its use in language understanding was illustrated in (Narayanan 1997a; Narayanan 1999a).

All the representations so far model states as propositions (or simple fluents) and are unable to handle relational information.[5] A critical aspect of scaling the current models to complex domain topologies and linguistic structure is the ability to model predicates and relations. Moving along the z-axis (into the plane), are relatively recent probabilistic models that handle relational information. The RMM model (Anderson *et al.* 2002) is a relational extension to sequential processes that allows variables and relations in markov models. However, unlike BN (or TBN), RMMs are unable to model complex states with dependencies. PRMs (Getoor *et al.* 2001; Pfeffer 2000) extend the Bayes Net formalism to allow specification of a probability distribution over a set of rela-

---

[4] The more recent versions of Hierarchical HMM (HHMM) (as in (Murphy 2002) are also subsumed under the TBN framework and such finer gradations are omitted in the figure to ease exposition.

[5] Our use of token types (Narayanan 1999b) (as in CSPN) allows for the action model to be relational but the state structure remains propositional.

tional interpretations. As in Bayes Nets, a PRM consists of a qualitative dependency structure and a set of parameters quantifying the conditional dependency. One basic difference is that in the case of PRMs we specify the dependency model quantifying the various domain relations at the class level. This dependency is assumed to be duplicated for each instantiation. While PRMs are able to model the relational information essential for modeling inference over complex-structured domains, they are unable to model coordinated action or dynamics. Moving rightward from PRMs we come to T(D)PRMs which are model simple sequences of PRMs in a manner analogous to T(D)BN.

Our existing formalism integrates extended CSPN nets for action and TBNs for inference and appears to be an attractive methodology for the construction and analysis of intelligent systems. For large scale systems, we will need to utilize more modular and structured probabilistic networks, or PRMs. Our current efforts at the computational level are directed towards a complete integration of DPRM and our extended Petri nets (see Section 4.1), which we call Coordinated Probabilistic Relational Models or CPRM (see Figure 3). These can be seen as extending an ongoing community effort to add more sophisticated temporal and control capabilities to probabilistic inference models (Anderson *et al.* 2002; Boyen & Koller 1998; Jordan 1999; Murphy 2002).

The rest of this section describes our technical approach to integrating the modular and structured inference provided by PRMs with the rich action modeling framework and distributed operational semantics provided by the PNs. Our action theory comprises of two central components; 1) an **executing** representation of *actions* (called **x-schemas**) based on extensions to Petri Nets and 2) a PRM model of *state* that captures and reasons about complex dependencies between state variables.

4.1 An expressive representation of coordinated actions and events

Central to the representation is an event model called **executing schemas** (or **x-schemas**), motivated by research in both sensorimotor control and cognitive semantics (Narayanan 1997a). X-schemas are active structures that cleanly capture sequentiality, concurrency and event-based asynchronous control. They thus provide a cognitively motivated basis for modeling diverse linguistic phenomena, including aspectual inference (Chang *et al.* 1998), metaphoric inference (Narayanan 1999a) and event-based reasoning in narrative understanding (Narayanan 1999b). We have built prototype systems (Chang *et al.* 2002) that demonstrate its use for a variety of *frame-based inferences* and the attendant problem of modeling *perspectival effects*.

X-schemas are based on Petri nets, which in its basic form is a weighted, bipartite graph consisting of **places** (shown as circles) and **transitions** (shown as rectangles) connected by directed input and output arcs (Reisig 1985; Murata 1989; Narayanan 1997a). Places may contain **tokens** (i.e., they may be *marked*), and they typically represent states, resources or conditions that apply. Transitions typically represent actions or events. X-schemas extend the basic Petri net to include typed arcs, hierarchical control, durative transitions, parameterization, typed (individual) tokens and stochasticity.

The most relevant property of the x-schema for action modeling is its well-specified execution semantics: a transition is **enabled** when all its input places are marked, such that it can **fire** by moving tokens from input to output places. The active execution semantics serves as the engine of context-sensitive inference in the simulation-based model of language understanding (described in Section 2).

Definition 1 The basic x-schema: An **x-schema** consists of *places*(P) and *Transitions* ($\mathcal{T}$) connected by weighted directed arcs $\mathcal{A}$ ($\mathcal{A} \in (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P})$). Each arc $a_{ij} \in \mathcal{A}$ has weight w_{ij} ($w_{ij} \in \mathcal{N}$). Input Arcs $*\mathcal{T}$ ($*\mathcal{T} \in (\mathcal{P} \times \mathcal{T})$) connect Input Places to Transitions. Output Arcs $\mathcal{T}*$ ($\mathcal{T}* \in (\mathcal{T} \times \mathcal{P})$) connect Transitions to Output Places. Arcs are typed as *enable* arcs $\mathcal{E}$, *inhibitory* arcs $\mathcal{I}$, or *resource* arcs $\mathcal{R}$.

X-schemas have a well specified real-time execution semantics where the **next state** function is specified by the **firing rule**. In order to simulate the dynamic behavior of a system, a **marking** (distribution of **tokens** in places (depicted as dark circles or numbers)) of the x-schema is changed according to the following **firing rule**.

Definition 2 Execution Semantics of the basic x-schema A transition $\mathcal{T}$ is said to be **enabled** if **no** *inhibitory* arc $i \in \mathcal{I}$ of $\mathcal{T}$ has a **marked** source place **and** all sources of enable arcs $e \in \mathcal{E}$ of $\mathcal{T}$ are **marked** and all input arcs $p \in \mathcal{R}$ have at least w_{pt} tokens at their source place, where w_{pt} is the weight of the arc from $\mathcal{P}$ to $\mathcal{T}$. The **firing** of an **enabled** transition $\mathcal{T}$, removes w_{PT} tokens from the source of each resource input arc $\mathcal{P}$ and places w_{TP} tokens in each output place of $\mathcal{T}$.

X-schemas cleanly capture sequentiality, concurrency and event-based asynchronous control; with our extensions they also model *hierarchy*, *stochasticity* and *parameterization* (run-time bindings). Besides *typed arcs* (Definition 1), the following two extensions to the basic Petri net are designed to allow us to model hierarchical action sets with variables and parameters: First, tokens carry information (i.e. they are individuated and typed) and transitions are augmented with predicates which select tokens from input places based on the token type, as well as relate the type of the tokens produced by the firing to the types of tokens removed from the input. Second, transitions are typed into four kinds, namely **stochastic**, **durative**, **instantaneous** and **hierarchical** transitions. An instantaneous transition *fires* as soon as it is enabled. A timed transition fires after a fixed delay or at an exponentially distributed rate. Hierarchical transitions, activate a subnet, wait for its return, or timeout.

The state of an x-schema is defined by its marking. The firing rule produces a change in the state of an x-schema by taking it from one marking to the next. Given an x-schema and a marking, we can execute the x-schema by *successive* transition firings. This can continue as long as there is at least one enabled transition that can fire. The x-schema firing rule semantics allows enabled transitions to fire in a completely distributed manner without any global clocks or central controllers.[6] Execution halts at the state where there is no enabled transition. This naturally allows us to extend the earlier definition to define an **extended next-state function** for x-schemas.

Definition 3 Extended next-state function

The extended next-state function is defined for a state s_i and a sequence of transitions $\rho \in T^*$ as

$$\begin{aligned} \delta(s_i, t_j\rho) &= \delta(\delta(s_i, t_j), \rho) \\ \delta(s_i, \lambda) &= s_i \end{aligned}$$

λ is the $null$ transition.

4.2 A PRM model of states

Our representation of states must be capable of modeling causal knowledge and be able to support both belief **updates** and **revisions** in computing the global impact of new observations and evidence both from direct observations and from action effects. Our implementation of the agent's state uses Belief Networks (Jensen 1996; Pearl 1988). A belief network is a convenient data structure to encode causal domain knowledge. The basic algorithms that operate on the probabilistic network data structure deal both with new observations and database updates due to external intervention such as actions and random disturbances.

In previous work (Narayanan 1999b), we have used temporally extended propositional graphical models or Dynamic (Temporally extended) Bayes Nets (TBN) (Pearl 1988; Jensen 1996; Jordan 1999; Murphy 2002) to model complex states. A Bayes Net is a convenient data structure to encode causal and propositional domain knowledge. The basic algorithms that operate on the probabilistic network data structure deal both with new observations and database updates due to external interventions such as new textual (or other) inputs.

A Bayes Net consists of a set of variables and a set of directed links. Each variable has a finite set of mutually exclusive states. The variables and links together form a DAG (Directed Acyclic Graph). To each variable A with parents $C_1 \ldots C_n$ there is attached a conditional probability table $P(A|C_1, \ldots C_n)$. A Bayes Network B_u can be viewed as a *compact* representation of the probability table, and if the conditional independencies in B_u hold for U, for a Bayes Net B_u, the following theorem allows us to calculate the joint probability $P(U)$ from the conditional probabilities in the network.

Theorem 1 . **The Chain Rule** *(Jensen 1996) Let* B_u *be a Bayes Net over* $U = (A_1 \ldots A_n)$. *Then the joint probability distribution* $P(U)$ *is the product of all conditional probabilities specified in* B_u.

$$P(U) = \prod_i P(A_i|pa(A_i)) \tag{1}$$

where $pa(A_i)$ *is the parent set of* A_i

---

[6] However our sequential simulation adjusts step size to be able to fire multiple enabled transitions in a single step.

We can use this recursive factorization for both *Belief Updates* and *MAP Estimation* to find the best explanation for an input narrative. Our results so far (Narayanan 1999b) suggest that this technique is promising enough to be useful for a variety of NLP applications.

While our current Bayes Net structure supports structured inferences inferences of interest to NLP applications (Narayanan 1999a), it does not exploit the relational structure inherent in any domain. In this project, we propose to use a recent extension to the Bayes Net formalism, Probabilistic Relational Models (PRMs) (Pfeffer *et al.* 2000; Pfeffer 2000; Pfeffer & Koller 2000; Getoor *et al.* 2001). PRMs extend the Bayes Net formalism to allow specification of a probability distribution over a set of relational interpretations. As in Bayes Nets, a PRM consists of a qualitative dependency structure and a set of parameters quantifying the conditional dependency. One basic difference is that in the case of PRMs we specify the dependency model quantifying the various domain relations at the frame level. This dependency is assumed to be duplicated for each instantiation. So every instance of the WALK schema will have the dependency model defined at the WALK frame level. A second difference is that the relational structure of the domain is explicitly used so that the role of one frame (*Walk.Agent*) can depend on the roles of related frames (including sub-frames) (*Step.Agent*).

Definition 4 A PRM Model of a State

The State $\mathcal{S}$ is defined as a PRM comprised of: A set $\mathcal{C}$ of classes, related by inheritance relationships. A set $\mathcal{I}$ of named instances, each denoting an instance of the *class.* A set A of *simple* attributes, denoting functional relations, where the domain $Dom[A] \in \mathcal{C}$ and the range some $Range[A] \in Val[A]$. A set B of *complex* attributes, denoting functional relations, where the domain $Dom[B] \in \mathcal{C}$ and the range $Range[B] \in \mathcal{C}$. A set of *Conditional Probability Distributions (CPD)* where for each class $c \in \mathcal{F}$ and for each attribute $a \in A \bigcup B$, we define a *CPD* of the form $P(a|pa(C.a))$, where $pa(C.a)$ are the parents of the attribute (simple or complex). CPDs are attached to classes and inherited by instances. Cyclic dependencies in the parent links are disallowed.

4.3 Inference With CPRMs

Traditional inference methods in probabilistic models of linear time dynamic systems (such as TBN) (see (Murphy 2002) for a fuller description) consist of the following kinds of computations. Here X_t is a state variable at time t (lowercase x_t is a value assignment), and y_t is an observation value at time t .

Filtering Compute $P(X_t|y_{i \ldots t})$. Update the state based on the observation sequence.

Prediction Compute $P(X_{t+h}|y_{1 \ldots t})$. Predict the state at some future time $t+h$ based on the observation sequence upto time t .

Smoothing Compute $P(X_{t-m}|y_{1 \ldots t})$. Recompute previously estimated states in the present of current evidence.

Viterbi Decoding Compute $argmax_{x_{1 \ldots t}} \quad P(x_{1 \ldots t}|y_{1 \ldots t})$. Compute the best assignment of values to the state given the observation sequence.

With the extension to **branching or coordinated** models of dynamic systems, we need to enhance these traditional inference procedures to include the computation of the **reachability** set of a *marking* in an x-schema based action framework.

Reachability is a fundamental problem in dynamic system theory and is central to computing a variety of event related inferences (Chang *et al.* 1998). In terms of x-schemas, give that the state space evolves through execution of enabled transitions, we can define the **reachable** states of an x-schema, given an initial marking.

Definition 5 Immediately reachable states

For an x-schema $\mathcal{S}$, with a state s_i , state s_j is **immediately reachable** if there exists a transition $t_k \in T$ such that $\delta(s_i, t_k) = s_j$.

Extending this concept, we can define the set of reachable markings for a given x-schema in some initial state. Basically, if s_j is immediately reachable from s_i , and s_k is immediately reachable from s_j , then s_k is in the reachability set of s_i . Thus the **reachability** relationship is the reflexive transitive closure of the immediately reachable relationship.

Definition 6 Reachability set

The reachability set $\mathcal{R}(\mathcal{S}, s_0)$ for an x-schema $\mathcal{S}$ with state s_0 is the smallest set of markings defined by a) $s_0 \in \mathcal{R}(\mathcal{S}, s_0)$, and b) If $s_j \in \mathcal{R}(\mathcal{S}, s_0)$ and $s_k = \delta(s_j), t_l$ for some $t_l \in \mathcal{T}$, then $s_k \in \mathcal{R}(\mathcal{S}, s_0)$.

Definition 7 Reachability Given an x-schema $\mathcal{S}$ with an initial state s_0 and a final state s_f , is $s_f \in \mathcal{R}(\mathcal{S}, s_0)$?

Given this definition of reachability, the following theorem allows us to directly use the vast number of techniques developed in the distributed systems literature for x-schema based inference.

Theorem 2 . *(Proof in Narayanan 1997a) An x-schema is formally equivalent to bounded High Level Generalized Stochastic Petri Net (HLGSPN). The reachability graph of a marked x-schema is isomorphic to a semi-Markov process.*

Given this theorem, we can compute the various parameters of interest in our x-schema based model of coordinated events. As in other models (Ciardo *et al.* 1994), we assume that the x-schema transition firing time is governed by an exponentially distributed random variable x_i . The firing time of transition t_i is given by

$$\mathcal{F}_{x_i} = 1 - e^{-\lambda_i x} \tag{2}$$

The negative exponential distribution renders the reachability graph of the SPN isomorphic to a continuous time Markov chain. The Markov chain MC can be obtained from the reachability graph as follows: The MC state space is the reachability set $R(PN)$ of the marked SPN . In MC , the transition rate from M_i to M_j is given by $q_{ij} = \lambda_k$, corresponding to the firing rate of the transition t_k from M_i to M_j . If several transitions lead from M_i to M_j , then q_{ij} is the sum of the rates of these transitions. If there is no link from M_i to M_j in $R(PN)$ then $q_{ij} = 0$ in MC .

The steady state distribution π of the MC is obtained by solving the linear equations:

$$\pi Q = 0 \tag{3}$$

$$\sum_{1=1}^{s} \pi_j = 1 \tag{4}$$

From the vector $\pi = (\pi_1, \ldots, \pi_s)$ we can compute the following measures.

Probability of being in a set of states: Let $\mathcal{B} \subseteq R(PN)$ constitute the states of interest in a given SPN . Then the probability of being in a state of the corresponding SPN is

$$P(\mathcal{B}) = \sum_{M_i \in \mathcal{B}} \pi_i \tag{5}$$

Probability of taking a transition t_j : Let EN_j be the subset of R(PN) in which the transition t_j is enabled. Then the probability that an observer looking randomly into the next sees t_j firing next (p_j) is given by

$$p_j = \sum_{M_i \in EN_j} \pi_i \frac{\lambda_j}{-q_{ii}} \tag{6}$$

where q_{ii} is the sum of the transition rates out of M_i .

The throughput of a transition is the mean number of firings at steady state.

$$d_j = \sum_{M_i \in EN_j} \pi_i \lambda_j \tag{7}$$

With this semantics in hand, we can now turn to one of the standard inference procedures in an action model (that of *temporal projection* or *prediction*).

The temporal projection problem consists of computing states s_{n+1} resulting from executing the action set $\mathcal{S} = [a_1, \ldots a_n]$ in a given initial state s_0 . The solution to this problem in our action model involves both simulating the direct temporal and causal structure of the action in the current context, as well as using the PRM (see Section 4.2) to compute the Maximum A Posteriori (MAP) estimates of related frames. The MAP estimates and related **belief_revision** (Pearl 1988) procedures allow us to compute the indirect effects (ramifications) that flow from the action (Narayanan 1999b). The basic algorithm is outlined below.

Algorithm 1 Temporal Projection

1. Set initial Marking M_0 . $\forall p \in s_0 : M_0(p) = 1, \forall p \notin s_0 : M_0(p) = 0$.
2. Fire enabled transitions $T_{e_0} \in T$ of M with initial marking M_0 . The next state function described earlier takes the system to a new marking M_{int0} . The state corresponding to this marking $S_{int} = \forall p : M_{int}(p) = 1 : p \in S_{int}, \forall p : M_1(p) = 0 : p \notin s_{int}$.
3. Run the **belief_revision** procedure to return the most consistent a posteriori assignment (MAP) of values to the state variables. The new state S_1 corresponds to the marking M_1 where $\forall p \in S_1 : M_1(p) = 1, \forall p \notin s_1 : M_1(p) = 0$.
4. **While** $1 \leq i \leq n$, **do**: fire enabled transitions $T_{e_{i-1}}$ with marking M_i . set $M_{int} = *a_i \bigcup M_i$. Run Step 3 to get S_{i+1} . **Return** S_{n+1} as the answer.

■

Steps 1 and 2 are essentially constant time, since our notion of state as a graph marking is inherently distributed over the network, so the working memory of an x-schema-based inference system is distributed over the entire set of x-schemas and state features. The result is a massively parallel solution to the projection problem. In addition, the central features of our action representation, namely that they are *executing* provides an elegant solution to the Frame Problem. Specifically, the action-based executing action semantics allows frame axioms to be implicitly encoded in the structure of the net and the local transition firing rules.

Step 3 requires MAP estimation of the state variables. Our in the past has used a TBN to model states. MAP estimation over such a network is well known to be intractable for complex domain topologies. While the worst case analysis does not change with the more expressive CPRM design, we propose to exploit the additional structure provided by the PRM framework to develop state and domain models to minimize the inferential complexity. The complexity of exact inference on PRMs (using the Structured Variable Elimination (SVE) algorithm is given below.

Theorem 3 . *(Pfeffer 2000) The space and time complexity of solving a query on* q *basic variables for a PRM* P *is at most* $O(Nkb^{k(m+2)}b^q)$ *, where N is the total number of attributes in* P *,* k *is the maximum number of interface variables for any object in* P *, and* m *is the maximum tree-width of any object in* P *.*

The two critical variables to control are k (the maximum number of interface variables) and m , the maximum tree-width or the maximum width of the dependency graph of a class, relative to an elimination ordering. For instance, if the dependency graph of a class is a *polytree*, then the (space+time) complexity of solving a query on q variables for a PRM P reduces to $O(Nkb^{(p+2)k}b^q)$, where p is the maximum number of parents of any attribute in P . Clearly, the cost is dominated by the size of the interface variables (k). In a general PRM , there may be no local method to guarantee that the interface size stays small. However by carefully introducing relations and doing query optimization, one can get a handle on this variable. For instance, if the only relation allowed is the *part-of* relation, then it is possible to specify at design time the maximum size of the interface variable k . This results in a specialization of PRMs called Object Oriented Bayes Nets (OOBN) (Pfeffer 2000).

Specifically, with respect to the structure of relational model for language, we are involved in the development of a semantically rich wide-coverage lexicon, FrameNet[7] in conjunction with linguists from UC Berkeley and ICSI. FrameNet represents frames, frame elements (FE), relations between frames (hierarchy and composition), and role relations between frames, FEs and filler types. We have encoded the relational structure of frames and the annotation instances (example annotations of corpus examples using the frame and role information) in a MySQL relational database (Fillmore *et al.* 2001). We are developing a Probabilistic Relational Model (PRM) of the lexical and annotation database along with an API that supports the kinds of queries NLP systems would typically use. As a PRM we specify the dependency model quantifying the various FrameNet relations at the frame level. This dependency is assumed to be duplicated for each annotation instance. We also exploit the relational structure of the FrameNet database to quantify role-binding constraints (where the FE of one frame (*Trial.Defendant*) can depend on the FEs of related frames (*Arrest.Suspect*). The current FrameNet design is easily handled as an OOBN (to represent frame

---

[7] FrameNet is funded by NSF ITR/HCI 0086132; URL:http://www.icsi.berkeley.edu/framenet.

inheritance) coupled to a an x-schema model (for the composition relation). Hence the tractability of the inference can be controlled by local design choices.[8]

We also propose to extend the inference algorithm for PRMs to the temporally extended version (TPRM). The simplest way to accomplish this is to encode the temporal relation between attributes of different time slices as another relation to the existing PRM and use the PRM inference algorithm on this enhanced network. However, it is often the case that elimination orderings on the enhanced PRM that do not explicitly take into account the temporal structure may produce extremely large cliques and make inference intractable. As in the case of TBN inference, we plan to explicitly use conditional independence information from the topology afforded by the temporal structure. Here we plan to draw from our own previous work on on-line Bayesian algorithms for sentence processing (Narayanan & Jurafsky 1998; Narayanan & Jurafsky 2001) and from the *interface* algorithm (Murphy 2002) for TBN. The interface algorithm defines the *outgoing interface* of a slice t in the DPRM as the set of variables I_t that have at least one child in the next slice $t+1$. We then use the result (from (Murphy 2002)) that the outgoing interface (I_t) d-separates the past $< t$ from the future $> t$ and hence provides a sufficient statistic for inference. Specifically, this allows for on-line inference algorithms similar to the inside-outside algorithm (for SCFG) and the forward-backward algorithm for HMM to be adapted to DPRM inference.

To summarize, in our model, executing actions is fast, parallel and reflexive, while inference with complex state dependencies to achieve global consistency is hard. To deal with the scaling problem, we propose to exploit the relational structure of knowledge (linguistic and domain) to allow for explicit domain modeling and query optimization choices to influence the tractability of inference. Together, this provides a formalism expressive enough to support deep semantic inference necessary for language understanding with guidelines and techniques to bound the time and space complexity. We expect that accomplishing this could result in the widespread use of structured probabilistic models in NLP, something long overdue.

5 Comparison to Logical Theories of Actions

Although the mechanisms outlined above were developed for language understanding, they seem to useful for some of the problems discussed in the recent AI literature. It is useful to compare our approach with logical theories of actions, for expressiveness, tractability of reasoning as well as learnability. In terms of representational power, the PRM framework is captured by the first-order probabilistic databases of the form described in (Pfeffer & Koller 2000; Kersting & Raedt 2001). However, these databases are unable to model the complex concurrency, resource and coordination primitives inherent in the CPRM model. In terms of the action model itself, we compare the CPRM framework with a specific representative language for reasoning about actions (please see (Narayanan 1999b) for further details). We assume the reader is familiar with reasoning with action models similar to (Gelfond & Lifschitz 1993).

As in $\mathcal{ARD}$ (Giunchiglia & Lifschitz 1995), we model both “inertial” (**always** C (where C is a formula)) and “dependent” (A **depends_on** B) fluents.

An $\mathcal{ARD}$ language consists of

1. A set of symbols $\mathcal{F}$, called *Fluent* names ($\mathcal{F} \neq \{\}$), and another disjoint set of symbols $\mathcal{A}$, called *Action* names ($\mathcal{A} \neq \{\}$) and $\mathcal{A} \bigcap \mathcal{F} = \{\}$. For every Fluent F there is a non-empty set Dom_F called the domain of F.

2. Two sets of symbols $I, D : I \in \mathcal{F} \wedge D \in \mathcal{F}$, where fluents in set I are called *Inertial* fluents and fluents in set D are called *Dependent* fluents.($I \bigcap D = \{\}$).

An *atomic formula* is of the form F **is** V, where V is in the domain of fluent F. If F is propositional, then $Dom_F = \{true, false\}$,and we write F for the formula F **is** $true$. A *value proposition* is of the form C **after** A^n, where C is a formula and A^n is a string of action names. If A^n is empty, we write **initially** C. A *constraint* is of the form **always** C, where C is a formula. An *effect proposition* is of the form A **causes** C **if** P, where A is an action, and C and P are formulas. A *dependency proposition* is of the form F **depends_on** G **if** P, where F is the dependent fluent name, G any fluent name, and P a formula.

The following rules present the basic encoding of action theories (assuming the syntax of the $\mathcal{ARD}$ theory) in our model.

---

[8] FrameNet continues to evolve, so more complicated frame relations may change this situation.

1. Static Fluent names are places. Actions names are Transition labels. Preconditions are pre-sets ($*\mathcal{T}$), direct effects are post-sets ($\mathcal{T}*$) of transitions. If the truth-value of a fluent $f \in \mathcal{F}$ is $true$, in State S_i , then the marking $M_i(f) = 1$.[9]

2. Domain Constraints with inertial fluents are modelled as instantaneous transitions. Statements in $\mathcal{ARD}$, of the form **always** $A \supset B$, add an instantaneous transition with pre-set $*\mathcal{T}_{\mathcal{AB}} = A$, post-set $\mathcal{T}_{\mathcal{AB}}* = A, B$. Dependent fluents are modeled as arcs in the Agent state Bayes net. More precisely, the statement f_j **depends_on** f_i **if** f_k , results in an arc from the variables representing f_i, f_k to f_j . The CPT entries for f_j are given by the appropriate constraints (including prior knowledge).

3. **Initially** C , is modeled by assigning an initial marking where $\forall f \in \mathcal{F} \bigcap C : M_0(f) = 1$. $\forall f_i, fj \in \mathcal{F}$, if f_i **depends_on** f_j , add an instantaneous transition $\mathcal{T}_{\mathcal{IJ}}$ with preset $*\mathcal{T}_{\mathcal{IJ}} = f_j$ and post set $\mathcal{T}_{\mathcal{IJ}}* = f_i, f_j$.

Theorem 4 . *(Narayanan 1999b) The procedure above results in a* **causal model** *for a domain description* D *(in the Syntax of* $\mathcal{ARD}$ *) in that it satisfies all the causal laws in* D *. Furthermore, a value proposition of the form* C **after** A *is entailed by* D **iff** $\forall c \in C, c \in S_i$ *where* S_i *is the state that results after running the projection algorithm on the action set* A .

In addition to the connection to the $\mathcal{ARD}$ language and its descendents, the CPRM framework can also be compared to representation efforts on the Semantic Web (http://www.semanticweb.org). The PRM framework has been used to model probabilistic version of description logics in (Koller *et al.* 1997). Description logics form the basis of the current RDF-based proposals, DAML+OIL and OWL (http://www.w3.org/TR/owl-ref). The x-schema based action framework has been shown (Narayanan & McIlraith 2002) to be capable of providing an operational semantics for plans in a rich situation calculus (Pinto 1994). (Narayanan & McIlraith 2002) also use the x-schema based action model as the semantic underpinning for the expressive Web-services markup language DAML-S (http://www.daml.org/services). We believe that the proposed integrated CPRM representation could be an adequate probabilistic model of the various emerging ontology markup languages on the Semantic Web.

One important feature of our proposed CPRM representation of events is that they are more fine-grained than other proposals we have seen. This finer granularity resulted from our interest in modeling aspectual distinctions made by different languages (Narayanan 1999a; Chang *et al.* 2002). We also believe that the finer-grained nature of our representation of aspect presents one natural way to constrain automatic inference in reasoning about event descriptions. Additional evidence comes from the observation that metaphoric projection of events across different domains appears to respect the temporal and aspectual distinctions made by our representation (Narayanan 1999a).

Other distinguishing features include the true concurrency semantics of our x-schema model, the ability to cleanly model uncertainty in action selection and action effects (both conditional and non-deterministic), a natural model of resource consumption and production, a model of continuous time (as in the previous section) and principled methods to support continuously varying state (Pinto 1994; Jordan 1999).

In general, we believe that probabilistic, graphical models are inherently desirable for action representation, inference and learning. The graph-based representation allows us to formally state and reason about inter-schema relations declaratively while using their real-time execution capability for inference. This allows our representation to be used to declaratively to specification and design or procedurally for projection and automatic inference. The factorized topology of the graph supports recursive parameter estimation techniques (Jordan 1999) and provides powerful constraints and inductive bias for relational structure learning (Getoor *et al.* 2001; Pfeffer & Koller 2000). We believe these properties to be essential for representations that are to be used both for acting and for reasoning about action descriptions.

6 Towards Integrated Large Scale NLU

Scalable probabilistic inference models such as CPRM are one critical requirement for large NLU systems, but are not sufficient. In this section, we will briefly review our understanding of the other barriers to scalable NLU and of current efforts to overcome them, both in our lab and elsewhere. The suggestion is that sufficient progress is being made on all fronts to allow serious testing of ambitious applications within the next few years. The two issues which will be

---

[9] In general, the representation allows states with types and integer measure fluents, in which case the multi-set representing the place would be marked by the appropriate number and type of tokens.

addressed are the availability of large structured knowledge bases and the feasibility of language analysis modules (parsers) that can produce the semantic structures needed for inference and understanding.

6.1 Ontologies and Linguistic Resources

We are interested in investigating methods and tools for the distributed development, management and use of open domain ontologies. We have had considerable interest from and experience with the Semantic Web community (Ankolekar *et al.* 2002; Narayanan *et al.* 2002; Narayanan & McIlraith 2002) and have developed a distributed operational semantics for the RDF-based Semantic Web Service markup language, DAML-S (http://www.daml.org/services) based the x-schema model of events. We have also developed an automatic translator from FrameNet to DAML+OIL (http://www.daml.org). We plan to continue our close association with this community and continue to leverage the growing set of tools, techniques and protocols to allow for an automatic extension to wide of adoption of our CPRM package.

Another critical requirement for scaling up NLU is the emergence of machine-interpretable language resources that are deep enough in structured semantic information to support true domain-independent understanding. The FrameNet project (Fillmore *et al.* 2001) has been building a unique kind of lexicon for English with properties that take it far beyond any other existing lexical resource. The building of the FrameNet lexicon began with linking groups of lexical units with low-level word-related semantic frames, but in its later evolution has been fitting these in turn into more global institutional and procedural frames. FrameNet is a structurally rich lexical database (LDB) based on the well developed theory of frame semantics (Fillmore 1968; Fillmore 1976; Fillmore & Atkins 1992).

We have built representations (Bergen & Chang 2002a) that bridge this gap through a formalism that unpacks the shorthand of frames into structured event representations. Our representation allows annotated FrameNet data to parameterize event simulations (Narayanan 1999b) that produce fine-grained, context-sensitive inferences. We have built prototype systems that suggest that the model can account for wide-ranging consequences of frame-based inferences (Narayanan 1997a; Chang *et al.* 1998; Narayanan 1999a; Chang *et al.* 2002).

6.2 Semantic Analysis

Even assuming that we can build scalable CPRM inference systems and link them with large knowledge bases and ontologies, another barrier to large scale NLU remains. We will need systems that can generate Semantic Analyses from input text (and eventually speech) as well as modules that can generate appropriate language output. There are two distinct approaches to this problem and we are pursuing both of them.

The knowledge intensive approach to semantic analysis focuses on the detailed subtle knowledge that supports human language and thought. Linguists continue to elaborate the intricacies of grammar and meaning and our systems should exploit this vast resource. Our work along these lines comprise two complementary research efforts. The Embodied Construction Grammar (ECG) project has focused on developing a formalism that is well suited to capturing the deep form-meaning relationships required for the embodied inference models described above.

A detailed overview of the Embodied Construction Grammar (ECG) formalism can be found in (Bergen & Chang 2002b) and has been presented at the First International Workshop on Scalable Natural Language Understanding at the European Media Laboratory. The formalism supports a new computational approach to cognitively motivated semantic analysis. Historically, there has been a gap between grammar formalisms precise enough to be used for parsing and those used in cognitive approaches to language. ECG bridges this gap by providing a formalism that marries the precision of unification grammar with the expressiveness of cognitive semantics. The basic linguistic unit of ECG is the construction, a pairing of form and meaning. On the form side, ECG allows for more expressive constraints, dropping the Phrase Structure Grammar requirements for strict constituent ordering. On the meaning side, ECG uses cognitively motivated schemas, including frames, image schemas and the x-schemas described in this proposal. These schemas are represented as feature structures with meaning constraints modelled as coindexation constraints.

ECG also extends standard unification-based formalisms with several novel operators that provide additional semantic expressiveness. The *evokes* operator makes related meanings accessible, thus allowing background frames and preconditions to be bound to the meaning pole. The *self* operator allows constructions to be self-referential (prohibited in traditional unification grammar) and the :: operator allows for conditional or dynamic unification that facilitates dealing with events. ECG has been successfully employed by linguists working in the construction grammar framework in different languages. This has reached the point where there are annual conferences on Construction Grammar (http://www.eng.helsinki.fi/janola/iccg2.htm) and there is also a special session on ECG at the 2003 ICLC meeting in Spain.

There is also a running Semantic Analysis system based on ECG (Bryant 2003 (to appear)) that is being used in various applications (Porzel & Bryant 2003 (to appear); Chang & Feldman 2003). This system has been designed explicitly to interface with large knowledge sources such as those described above. The technical idea is to add a restricted set of predicates to the constraints of ECG and to define an application program interface(API) for linking to ontologies such as CYC or Wordnet. We have completed the initial version of the API and are testing the compatibility of existing ontologies with the needs of ECG analysis.

Of course, large scale use of ECG will require an extensive compendium of constructions, but this is what many linguists do for a living. Our (admittedly visionary) vision of how this might all work comes from organic chemistry. Chemists have developed notations and disciplinary coordination that allows then to cumulatively describe a collection of objects and interactions that is of complexity comparable to language. The FrameNet effort and the general success of Open Source Software suggests that it is not unreasonable to work towards a scientific linguistics that will yield the wisdom needed for a knowledge based NLU. In fact, the world-wide HPSG consortium has gone far towards building a community effort on common grammars and analyzers for several languages (http://www.ling.ohiostate.edu/research/hpsg/). The form side of ECG shares many features with HPSG and it is at least possible that some synthesis could be brought about.

Another stream of analysis comes from learning approaches currently pursued by many researchers that are attempting to automatically extract semantic relations from text (Gildea & Jurafsky 2002; Harabagiu & Moldovan 1998; Harabagiu 2002; Rosario *et al.* 2002; Mohit & Narayanan 2003) The learning techniques used cover the spectrum from supervised methods (that need high quality annotations) to weakly supervised and unsupervised methods.

We believe that it is necessary for the two approaches to coarticulate and inform each other. Corpora labeled using knowledge intensive analyses can be used as training data for learning systems. For example, Gildea's program (Gildea & Jurafsky 2002) for learning to extract semantic roles from text was trained on data from the FrameNet corpus. On the other hand, statistical regularities revealed by learning studies can inform the design of constructions and of semantic analysis systems that attempt to find most likely descriptions. There is a wide range of efforts following both paths. Specifically, we are pursuing[10] the path of combining statistical learning techniques with construction parsing in the AQUAINT project as part of the ARDA program on question answering systems. Again, it would be naive to assume the semantic analysis problem can be easily solved, but equally naive, and a lot less productive, to give up trying.

7 Conclusion

Scalable probabilistic inference models such as CPRM are not sufficient for large NLU systems, but are certainly necessary. We propose to develop the theory of CPRM, implement a generally available package that efficiently realizes the theory, and demonstrate its inferential capabilities on a scalable version of the existing pilot understanding system (Section 2). As parallel efforts to build semantic analyzers and on wide-coverage ontologies and knowledge compendia mature, we will be ready to exploit these resources for large-scale applications in NLU. In addition, the CPRM package could be potentially useful in a wide range of scientific and practical tasks that are now using one of the less expressive members of the model space of Figure 4. Assuming parallel progress on the other key issues, CPRM could form the inferential core of scalable NLU systems that could radically change who can use computers and what can be done with them. In any event, the proposed research should add to our understanding of both probabilistic reasoning and natural languages.

---

[10] The project is a collaborative effort that involves Chris Manning from Stanford University, Marti Hearst from SIMS, UC Berkeley and Feldman and Narayanan from ICSI, Berkeley.

References

ANDERSON, CORIN R., PEDRO DOMINGOS, & DANIEL WELD. 2002. Relational markov models and their application to adaptive web navigation. In *Proc. KDD-2002*, http://www.cs.washington.edu/homes/weld/papers/kdd02.pdf.

ANKOLEKAR, ANUPRIYA, MARK BURSTEIN, JERRY R. HOBBS, ORA LASSILA, DAVID L. MARTIN, SHEILA A. MCILRAITH, SRINI NARAYANAN, MASSIMO PAOLUCCI, TERENCE PAYNE, KATIA SYCARA, & HONGLEI ZENG. 2002. Daml-s: A web service description for the semantic web. In *Proceedings of the First International Conference on the Semantic Web*.

BAILEY, DAVID R., NANCY CHANG, JEROME A. FELDMAN, & SRINI NARAYANAN. 1998. Extending embodied lexical development. In *Proc. 20th Cognitive Science Society Conference*, 84–90, Boston.

BERGEN, BENJAMIN K., & NANCY C. CHANG. 2002a. Embodied Construction Grammar in simulation-based language understanding. Technical Report TR-02-004, International Computer Science Institute.

——, & ——. 2002b. Simulation-based language understanding in Embodied Construction Grammar. In *Construction Grammar(s): Cognitive and Cross-language dimensions*, ed. by Jan-Ola Östman. John Benjamins.

BOYEN, XAVIER, & DAPHNE KOLLER. 1998. Tractable inference for complex stochastic processes. In *Proceedings of the Conference on Uncertainty in AI, UAI-1998*, 33–42.

BRYANT, J. 2003 (to appear). Constructional analysis with embodied construction grammar. In *Proc. International Cognitive Linguistics Conference*, La Rojas, Spain.

CHANG, N., & J. FELDMAN. 2003. Embodied construction grammar in learning and use. In *Proc. GURT Conference*, Washington D.C.

CHANG, NANCY, DANIEL GILDEA, & SRINI NARAYANAN. 1998. A dynamic model of aspectual composition. In *Proc. 20th Cognitive Science Society Conference*, Madison, Wisconsin.

——, SRINI NARAYANAN, & MIRIAM R.L. PETRUCK. 2002. Putting frames in perspective. In *Proc. Nineteenth International Conference on Computational Linguistics (COLING 2002)*.

CIARDO, GIANFRANCO, REINHARD GERMAN, & CHRISTOPH LINDEMANN. 1994. A characterization of the stochastic process underlying a stochastic petri net. *Software Engineering* 20.506–515.

DEAN, TOM, & MICHAEL WELLMAN. 1991. *Planning and Control*. Morgan Kaufman series in Representation and Reasoning.

FELDMAN, JEROME. 1990. Computational constraints on higher neural representations. In *Computational Neuroscience*. Cambridge, MA: MIT Press.

FILLMORE, CHARLES J. 1968. The case for case. In *Universals in Linguistic Theory*, ed. by Emmon W. Bach & Robert T. Harms, 1–88. New York: Holt, Rinehart & Winston.

—— 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, 20–32.

——, & B.T.S. ATKINS. 1992. Towards a frame-based lexicon: The semantics of RISK and its neighbors. In *Frames, Fields and Contrasts*, ed. by Adrienne Lehrer & Eva Feder Kittay. Lawrence Erlbaum Associates.

——, CHARLES WOOTERS, & COLLIN F. BAKER. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*, Hong Kong.

GALLESE, V., L. FADIGA, L. FOGASSI, & G. RIZZOLATTI. 1996. Action recognition in the premotor cortex. *Brain* 119.593–609.

GELFOND, M., & V. LIFSCHITZ. 1993. Representing action and change by logic programs. In *Journal of Logic Programming*, 17:301–322.

GETOOR, LISE, NIR FRIEDMAN, DAPHNE KOLLER, & AVI PFEFFER. 2001. Learning probabilistic relational models. In *Relational Data Mining*, ed. by DzeroskiLavrac, 307–335. SV.

GILDEA, DANIEL, & DANIEL JURAFSKY. 2002. Automatic labeling of semantic roles. *Computaional Linguistics* 28.245–288.

GIUNCHIGLIA, E., & V. LIFSCHITZ. 1995. Dependent fluents. In *Proceedings of the International Joint Conference on Artificial Intelligence, (IJCAI 95)*, 1964–1969.

GRAFTON, SCOTT T., MICHAEL A. ARBIB, L. FADIGA, & G. RIZZOLATTI. 1996. Localization of grasp representations in humans by PET: 2. Observation compared with imagination. *Experimental Brain Research* 103–111.

HARABAGIU, S. 2002. Just in time question answering. In *http://www.cs.utdallas.edu/ sanda/papers/nlprs.ps.gz*. Web Document.

——, & D. MOLDOVAN. 1998. Knowledge processing in an extended wordnet. In *Wordnet: An Electronic Lexical Database*, 379–405. MIT Press.

HOBBS, JERRY R., MARK STICKEL, PAUL MARTIN, & DOUGLAS EDWARDS. 1988. Interpretation as abduction. In *Proceedings of the 26th ACL*, 95–103, Buffalo, NY.

JENSEN, FINN V. 1996. *Introduction to Bayesian Networks*. Springer-Verlag.

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

JORDAN, MICHAEL I. 1999. *Learning in graphical models*. MIT Press.

KERSTING, KRISTIAN, & LUC DE RAEDT. 2001. Towards combining inductive logic programming with Bayesian networks. *Lecture Notes in Computer Science* 2157.118–??

KOLLER, DAPHNE, ALON Y. LEVY, & AVI PFEFFER. 1997. P-CLASSIC: A tractable probablistic description logic. In *AAAI/IAAI*, 390–397.

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

——. 1994. What is metaphor. In *Advances in Connectionist Theory, V3: Analogical Connections*, ed. by Barnden J. & K Holyoak. Addison-Wesley.

——, & MARK JOHNSON. 1980. *Metaphors We Live By*. University of Chicago Press.

LANGACKER, RONALD W. 1987. *Foundations of Cognitive Grammar, Vol. 1*. Stanford University Press.

—— 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Cognitive Linguistics Research. Berlin and New York: Mouton de Gruyter.

LERNER, U., B. MOSES, M. SCOTT, S. MCILRAITH, & D. KOLLER, 2002. Monitoring a complex physical system using a hybrid dynamic bayes net.

MOENS, MARC, & MARK STEEDMAN. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14.15–27.

MOHIT, B., & S. NARAYANAN. 2003. Semantic extraction with framenet. Technical report, International Computer Science Institute, Berkeley, CA.

MURATA, TADAO. 1989. Petri nets: Properties, analysis, and applications. In *Proc. IEEE-89*, volume 77, 541–576.

MURPHY, KEVIN, 2002. *Dynamic Bayesian Networks: Representation, Inference, and Learning*. University of California, Berkeley dissertation.

NARAYANAN, SRINI, 1997a. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. Computer Science Division, University of California at Berkeley dissertation.

——. 1997b. Talking the talk is like walking the walk: A computational model of verbal aspect. In *Proc. 19th Cognitive Science Society Conference*.

——. 1999a. Moving right along: A computational model of metaphoric reasoning about events. In *Proc. Sixteenth National Conference of Artificial Intelligence (AAAI-99)*. AAAI Press, Menlo Park.

——. 1999b. Reasoning about actions in narrative understanding. In *Proc. Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*. Morgan Kaufmann Press.

——, CHARLES J. FILLMORE, COLLIN F. BAKER, & MIRIAM R. L. PETRUCK. 2002. Framenet meets the semantic web: A DAML+OIL frame representation. In *Language Resources and the Semantic Web. AAAI (2002)*, Edmonton. AAAI.

——, & DANIEL JURAFSKY. 1998. Bayesian models of human sentence processing. In *Proc. 20th Cognitive Science Society Conference*, 84–90. Lawrence Erlbaum Associates.

——, & DANIEL JURAFSKY. 2001. A bayesian model predicts reading times in human sentence processing. In *Proc. Neural Information Processing Systems (NIPS)*.

——, & SHEILA MCILRAITH. 2002. Simulation, verification and automated composition of web services. In *Proc. Eleventh International World Wide Web Conference (WWW2002)*.

PEARL, JUDEA. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

PFEFFER, AVI, 2000. *Probabilistic Reasoning for Complex Systems*. Stanford University dissertation.

——, & DAPHNE KOLLER. 2000. Semantics and inference for recursive probability models. In *AAAI/IAAI*, 538–544.

——, ——, BRIAN MILCH, & KEN T. TAKUSAGAWA. 2000. SPOOK: A system for probabilistic object-oriented knowledge representation. In *International Conference on Uncertainty in AI*, 541–550.

PINTO, JAVIER A., 1994. *Temporal Reasoning in the Situation Calculus*. Toronto: Department of Computer Science, University of Toronto dissertation.

PORZEL, R., & J. BRYANT. 2003 (to appear). Construal resolution with embodied construction grammar. In *Proc. International Cognitive Linguistics Conference*, La Rojas, Spain.

REISIG, WOLFGANG. 1985. *Petri Nets: An Introduction*. Berlin: Springer–Verlag.

ROSARIO, BARBARA, MARTI HEARST, & CHARLES J. FILLMORE. 2002. The descent of hierarchy, and selection in relational semantics. In *Proceedings of the Association for Computational Linguistics*, Philadelphia. Association for Computational Linguistics.

SLOBIN, DAN ISAAC. 1997. The origins of grammaticizable notions: Beyond the individual mind. In *Expanding the Contexts*, ed. by Dan Isaac Slobin, volume 5 of *The Crosslinguistic Study of Language Acquisition*, chapter 5. Mahwah, New Jersey; London: Lawrence Erlbaum Associates.

STEEDMAN, MARK. 1996. Temporality. In *Handbook of Logic and Language*. North Holland: Elsevier.

TALMY, LEONARD. 1988. Force dynamics in language and cognition. *Cognitive Science* 49–100.

——, 1999. Spatial schematization in language. Presented at Spatial Cognition Conference, U.C. Berkeley.

TANJI, JUN. 1994. The supplementary motor area in the cerebral cortex. *Neuroscience Research* 19.251–268.

VENDLER, ZENO. 1967. *Linguistics in Philosophy*. Cornell University Press.

VERKUYL, HENK J. 1993. *A Theory of Aspectuality*. Cambridge University Press.