Title
A Dynamic Model of Aspectual Composition

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Authors
Chang, Nancy
Gildea, Daniel
Narayanan, Srini

Publication Date
1998

Peer reviewed

A Dynamic Model of Aspectual Composition

Nancy Chang, Daniel Gildea and **Srini Narayanan**
{nchang, gildea, snarayan}@icsi.berkeley.edu
International Computer Science Institute and University of California, Berkeley
1947 Center Street Suite 600, Berkeley CA 94704 USA

Abstract

This paper describes results of a dynamic model of aspectual composition that demonstrates how features necessary for planning and controlling actions can also motivate and ground simple analyses of a number of aspectual phenomena. A novel feature of the model is an active computational representation for verb semantics called **x-schemas**, an extension of the Petri net formalism that can encode goals, resources and other features affecting aspect. Vexing problems of aspectual composition lend themselves to simple analyses in terms of the context-sensitive interaction between verb-specific x-schemas and a **controller** x-schema that captures important regularities in the evolution of events. The resulting x-schemas can be elaborated and constrained by such factors as tense, temporal modifiers, nominals and pragmatic context, providing a rich representation that supports simulative inference in language understanding.

Introduction

Since Vendler's seminal paper (1967), the complex and context-sensitive determination of aspectual status, or the internal temporal shape of an event, has been the focus of much work (Comrie, 1976; Dowty, 1979; Verkuyl, 1993; Hwang and Schubert, 1994; Steedman, 1996).

Moens and Steedman (1988) (M&S) make the crucial observation that aspectual phenomena depend on a notion of event structure that captures **contingency** relationships among events. They use a tripartite event **nucleus** to characterize inherent aspectual classes and offer a transition network that specifies possible **coercions** between classes induced by grammatical markers and temporal modifiers. In *Harry is hiccupping*, for example, the progressive coerces the hiccup into an iterative process. M&S also illustrate how notions of contingency affect the interpretation of temporal connectives like *when*.

This paper describes a dynamic model of aspectual composition that demonstrates how features needed for planning and controlling actions can also motivate and ground simple analyses of a number of linguistic phenomena. Our work extends M&S's analysis by providing an active computational model with motivated constraints on aspectual coercion and a more precise description of inter-event contingency.

(Bailey et al., 1997; Narayanan, 1997a; Narayanan, 1997b) introduced our basic computational model of verbs and events, motivated by recent results in biological control theory. A novel feature of the model is an event representation called an **executing schema**, or **x-schema**, an extension of the Petri net formalism (Reisig, 1985). X-schemas are active structures that tightly couple action and reaction and are highly responsive to changes in both the environment and intentional state. Such properties are precisely those needed for controlling goal-directed behavior in a complex, uncertain and dynamic environment. They also, however, provide a cognitively motivated basis for many semantic distinctions that arise from the highly context-sensitive interaction between lexical and grammatical aspect, nominal semantics and tense. The dynamic model we have implemented supports fine-grained event simulation needed for inference in language understanding.

(Narayanan, 1997b) showed how our dynamic model of verb x-schemas interacting with their controller x-schema could handle some subtle interactions between inherent and phasal aspects and avoid some well-known paradoxes and problems in other accounts. In this paper, we report on the applicability of our model to the general problem of aspectual composition, including the interaction of verbal aspect with nominals, tense, temporal modifiers and pragmatic context. We first briefly revisit the basic event representation, focusing on extensions that allow us to model the composition of multiple features in a dynamic system. We then outline the results of applying our model to the various aspectual phenomena.

Representation

The unified representation we employ for verbs and the actions and events they describe is partly inspired by recent results in biological motor control theory. Our computational model is based on extensions to the Petri net formalism. The basic Petri net is a weighted, bipartite graph that consists of **places** (drawn as circles) and **transitions** (drawn as rectangles) connected by directed input and output arcs. The state of a net is defined by a **marking** that specifies a distribution of **tokens** (specified as a black dot or a number) over the places of the net. The real-time execution semantics of Petri nets models the production and consumption of resources: a transition is **enabled** when all its input places are marked such that it can **fire** by moving tokens (the number specified by the weight of the arc) from input to output places.

The most relevant features of Petri nets for our purposes are their ability to model events and states in a distributed system and cleanly capture sequentiality, concurrency and event-based asynchronous control. Our extensions to the basic Petri net formalism include typed arcs, hierarchical control, durative transitions, parameterization, typed (individual) tokens and stochasticity. We will illustrate these extensions through the simple x-schema for WALK(TO STORE) shown in Figure 1. This schema depicts conditions (such as visual and postural conditions) that allow an agent with sufficient energy to begin an ongoing process of walking by taking a step with each foot, which continues until the agent arrives at the store.

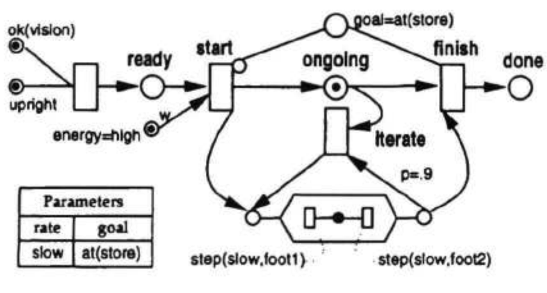

Figure 1: X-schema for WALK(TO STORE)

The introduction of **typed arcs** allows places to represent a variety of conditions on transitions. Standard Petri net input arcs, or **resource** arcs, model measure fluents such as energy and force. In the schema shown in the figure, the energy place acts as a resource that must have at least w tokens before the start transition fires. **Enable** arcs (shown as undirected) and **inhibitory** arcs (with inhibitory circles) allow places to act as pre-conditions and post-conditions that enable and disable schemas, respectively. The goal of being at(store) in Figure 1 does both: it inhibits the start of the schema (and perhaps other schemas that accomplish the same goal, such as DRIVE TO STORE), and it is a precondition to finishing the action. The revised execution semantics add the condition that transitions fire only if no inhibitory arc is marked, and all enable arcs are; these arcs result in no token transfer.

Hierarchical control allows us to model action hierarchies and the decomposition of transitions into subschemas. In the figure, WALK is decomposed (shown as a hexagon) into a sequence of steps; these could be further refined to stance and swing phases. X-schema transitions can have real durations which correspond to the delay between enabling and firing. Place markings persist until they are drained by a transition. States that hold for a certain period can thus be modeled by attaching a sink transition with a specific duration.

X-schemas have several properties that allow them to adapt to a dynamic environment, rendering it impossible to view x-schemas merely as pre-compiled programs. **Parameterization** allows x-schemas to **dynamically bind** to different world objects, corresponding to **individual tokens**, and to operate with different parameters during execution. The schema in the figure, for example, dynamically binds store to its goal location and is parameterized as walking at a slow rate. Stochastic transitions and arcs model uncertainty in world evolutions as well as prioritized action selection. In the figure, the high probability p=.9 of taking the iterate rather than finish arc after two steps models the inherently periodic or continuous nature of walking.

Flow of activation within a single schema is locally controlled, but it can be maintained until a goal is achieved, and it is susceptible to event-based interruption. X-schemas can monitor resources and the effects of their execution, as well as distinguish between simple enabling and the actual execution of an action. A much fuller discussion of the formalism, including a proof of a theorem establishing the formal connection between x-schemas and Generalized Stochastic Petri Nets (GSPN), can be found in Narayanan's dissertation (1997).

The properties described here, in addition to those of Petri nets in general, are just a few of those that make x-schemas an ideal representation for the control and coordination of complex actions in a dynamic environment. The next section shows how the same properties prove essential for modeling the semantics of aspect. Specifically, linguistic elements provide information required for the initiation and control of underlying x-schemas; we then describe how information provided by different elements can be composed. The well-defined execution semantics of the formalism then allows real-time simulation that may be necessary for inference in language understanding.

Aspectual composition

The semantics of aspect arises from the dynamic binding between verb-specific x-schemas and a **controller** x-schema that captures regularities in the evolution of complex events, shown in Figure 2.

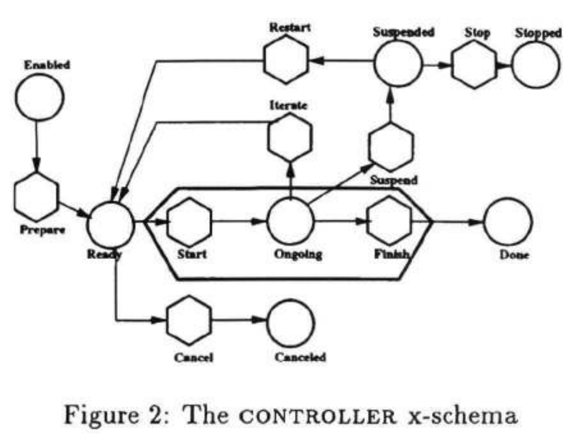

Figure 2: The CONTROLLER x-schema

In many events, particular preconditions must be met before an event is ready to start, and even then it may be canceled. While the event is ongoing it may iterate many times before it is done (usually with consequences), as in Figure 1. Since the structure of the controller constrains possible execution trajectories, a marked ongoing node indicates that the activity in question has already started and might next iterate.

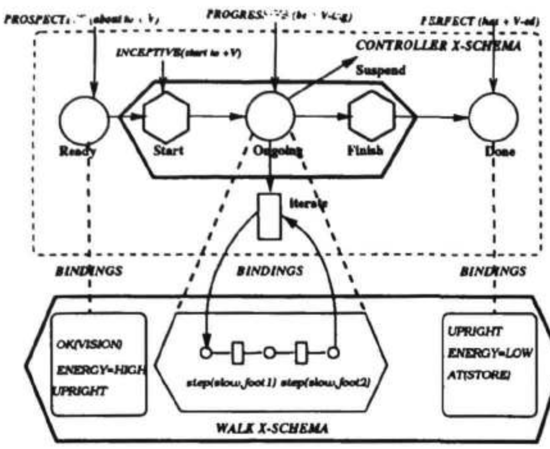

Figure 3: Binding of WALK to controller

Crucially, the controller is itself an active x-schema that can thus interact with both the underlying schema and the world state. Figure 3 shows how nodes of the controller can be bound to different phases of the WALK x-schema. The interpretation of **phasal** aspect falls directly out of bindings of particular controller nodes to the underlying schema. Figure 3 shows how the **prospective** aspect correspondings to the state of being about to walk and specifies that the agent is upright, has enough energy, etc., while the **progressive** aspect binds to an ongoing walk process, indicating that steps are being taken and energy being consumed. Lexical items can correspond to specific markings of the controller and affect the execution trajectory of the underlying x-schema; verbs in infinitive form refer to the entire graph.

Cross-linguistic evidence for the model comes from the ubiquity of aspectual expressions that directly activate specific nodes of the controller (Comrie, 1976), though we focus on English in this paper. These may be either lexical items (English *start to V*) or grammatical markers (Tamil *ind* marks iterate; Mandarin *le* and English perfect mark done). The controller transitions can themselves be decomposed into subordinate aspect graphs, capturing the semantics of *almost started to walk*, which expands start and marks its canceled node.

Conflicting requirements of different lexical items prevent some combinations. English progressive activates the ongoing node, which cannot itself be decomposed and started. Thus the preferred interpretation of *started almost walking* is one in which *almost walking* refers to some action distinct from walking, such as crawling. More complex aspectual patterns arise when languages grammaticize more complex controller markings.

Although the controller most naturally applies to complex, dynamic events, even apparently static situations may be viewed as corresponding to the ongoing node. This view of states allows our model to preserve the general restriction against states in the progressive (**He is being tall*) but also account for challenging exceptions in a principled manner. Although prototypical states may be shown as simple x-schema places, they can also correspond to the ongoing state of a simplified controller, as in Figure 4(a).[1] Although both *I live in Texas* and *I am living in Texas* may refer to the same ongoing state, they differ in scope: the present tense version focuses on the state itself, while the progressive version has the endpoints of the state in focus and thus takes on a temporary reading.[2] The ongoing node also often requires a resource to persist, as in (b), licensing progressive marking and temporal, intentional readings (*Bill's being silly*).

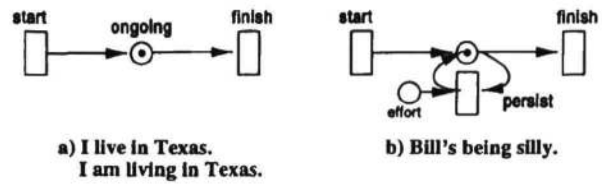

Figure 4: States viewed through the controller

Features of the subject may influence the salience of each of these readings. In *The lamp is standing by the doorway*, the inability of the inanimate subject to provide resources like energy or intention forces a temporary reading not possible for the more permanent subject of **The moat is surrounding the castle.*

Aspectual classes

As evidenced by the complications that arise even for simple states, aspectual composition is highly sensitive to constraints imposed by context and clausal elements. The notion of composition that emerges in our model, however, differs dramatically from the logical composition of static aspectual features and classes common in the literature. The controller and the specific lexical and grammatical items in a sentence jointly account for acceptability judgments and contextually salient interpretations. Aspectual inference draws on a more flexible notion of coercion than is possible for theories based on strict aspectual classes; it is characterized instead in terms of the dynamic interaction of the various schematized process primitives outlined earlier. In that sense our model elaborates on Steedman's more recent proposal (1996) to use dynamic semantics to characterize aspect.

The aspectual distinctions proposed by Vendler, M&S and others are a derivable subset of those in our model. Resource consumption in **continuous** processes allows simple distinctions to be drawn between states and processes, and the standard **telicity** feature can be treated simply as progress toward some goal, often depleting some limited resource, e.g., the distance to the park in *walk to the park*.

Additional flexibility arises from the fact that specific events may require only a subpart of the overall controller, as demonstrated for states earlier. The model

[1] Similarly, because verbs that are non-intentional by default, as in *He lost his keys*, do not have a preparatory phase, their enabled and ready nodes may be collapsed into a single enabled node.

[2] We label the controller transition persist here instead of iterate to emphasize that the same state is being maintained.

thus allows the same situation to be construed in subtly different ways; the well-known Imperfective Paradox is easily resolved by observing that the actions of Figure 1 may or may not be seen as goal-oriented. Like states, a continuous process like walking has no larger goal-oriented structure and no condition preventing it from finishing. This lexically specified aspectual structure may interact differently with particular aspectual markers. For instance, since English perfect marks the done node, goal-oriented processes like that depicted in *John has walked to the store* clearly license inferences like at(store). But it is possible to infer *John has walked* at any time after the lower-level continuous process (taking two steps) has been executed, since in the absence of a goal, the done transition can be taken without asserting that the ongoing state or process has ended.

The perspectival shifts important for making the basic **imperfective/perfective** distinction are naturally modeled by hierarchical decomposition of transitions, as well as abstraction from the controller to a **discrete** transition. Some shifts may prevent reference to controller nodes, thus accounting for iterative readings (as in the *sneezing* examples).[3] Depending on the particular combination, however, the controller may allow instead of prevent access. The verb *win*, for instance, is usually seen as a punctual event. In the framework given here, not only can it be placed in the context of a race — shown in Figure 5 as the finish transition of a *race* — but in combination with progressive marking, as in *John was winning the race*, the transition can be expanded into an ongoing subevent that can even be modified.

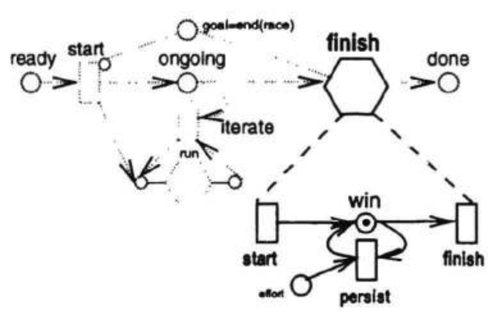

Figure 5: Progressive marking of *win* as finish of a race

The remainder of this section will be devoted to characterizing some of the constraints imposed by clausal elements, including tense, temporal modifiers and nominals.

Tense

Our model interprets tense as contributing information about the time of an x-schema activation state relative to the time of speech. This approach incorporates observations from Reichenbach's (1947) classic analysis of the relations among speech (S), event (E) and reference (R) time, but we can more clearly characterize both R and E by projecting the controller onto a timeline, as shown in Figure 6.

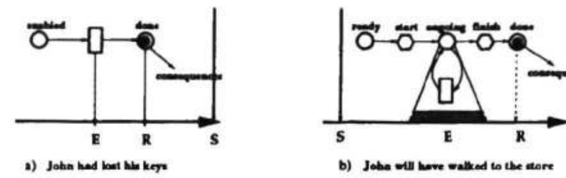

Figure 6: Projecting the controller over time

E corresponds to the duration of the event; for discrete events like (a), E will, as in Reichenbach, map to a particular point on the line. R is determined by the controller state specified by the event description; here the past perfect of *lose* marks the done node of the (simple) controller, allowing the inference that E preceded both the reference and the utterance time.

The advantage of using the controller becomes clearer when more of the controller structure is accessible, as in the future perfect in (b). In particular, we specify E as mapping to the time (interval) during which the controller's ongoing node is marked. Note that S potentially takes place before, during or after the projection of the ongoing node onto the timeline (shown as after), providing an explanation for the lack of linguistic markers expressing a relation between S and E in Reichenbach's analysis.

Temporal modifiers

Temporal modifiers like *for five minutes* and *in an hour* have been used from Vendler onward to distinguish aspectual classes. Both are easily handled as delay on the firing of x-schema transitions, as mentioned earlier, but they differ in the requirements they place. Specifically, we model **durative** (*for-*) modifiers as a delay on the suspend transition in the controller scheme; for states this reduces to the finish transition. **Span** (*in-*) modifiers apply to a goal-enabled finish transition, modifying the duration of both the ongoing process and the finish transition. They may also directly mark a transition to indicate simply that the transition fires after the specified time.

These requirements explain the differing inferences about whether the book has been finished between the following:

1. a. She read the book for an hour.
   b. She read the book in an hour.

Similarly, duratives naturally combine with states and processes, but spans produce marked inceptive readings (*Bill walked {for/*in} an hour*).[4]

[3] Our model also accommodates perspectival shifts based on pragmatically determined time scale of reference (e.g., slow-motion movie contexts allowing *He sneezed for 5 minutes*). Further details, including the modeling of viewpoint aspect, can be found in (Narayanan, 1997a).

[4] Some languages freely allow inceptive reading for past perfective states, as in Spanish *(Yo) le conocí al hombre*, where the perfective form of the verb "know" translates as "I met the man"

The well-known iterative coercion induced by duratives has a motivated explanation in this case, since the suspend transition can fire only after the specified time period, forcing discrete transitions like *sneeze* to take the iterate loop.

Durative modifiers sometimes apply to a state resulting from an event rather than the event itself: *Jill left the room for an hour* refers to the period after Jill has left; the salience of the state may explain why this reading is preferred to an iterative one. The inference that she returns at the end of this period falls naturally out of the fact that states given a duration are implied to finish after the specified duration, which is most easily effected in this case by a return to the previous state.

Such durative modification of a result seems to take place most easily for events construed as simple transitions, even when the underlying event is a more complex continuous process. Referring once again to Figure 1, the depicted event could be described by either of the verbs *walked* or *went*, but addition of *for an hour* produces differing patterns of most likely inference:

2. a. John walked to the store (for an hour).
   b. John went to the store (for an hour).

Additional aspectual mysteries are explained by this interpretation of temporal modifiers. Transitions that take place in the context of a larger goal-directed event easily satisfy the requirements of span modifiers (*Marc won the race in 5.3 minutes*), but they are more resistant to appearing with duratives: **Marc won the race for a few minutes.* Acceptable interpretations might result from iteration (especially when the duration is *for years*), but since no reversible state results the duration cannot apply to the done state. It can, however, apply to the ongoing state that has been brought into focus by progressive marking

Nominal features

Nominal distinctions like number, boundedness (mass or count), animacy and specificity have widely noted aspectual consequences (Verkuyl, 1993; Langacker, 1991). the simplest case, plural subjects or objects can provide another route to iteration, as in *Soldiers were reaching the summit all morning.*

In more complex cases, we interpret nominals as affecting the **resources** provided to x-schemas. As seen earlier with progressive states, the ability of nominals to furnish intentional resources plays a large role in distinguishing kinds of states and processes; a stative sense of *run* results in **The road is running from here to the school.*

Similarly, present tense applies only to states that hold at utterance time and thus favors the stative reading. When no such reading is possible, present tense often produces a habitual reading (via iteration into a larger-scale currently ongoing process): *Mary runs from here to the school* has a habitual reading, unlike *The road runs from here to the school.*

Resources consumed during the event also provide goals that enable finishing. This analysis accounts for the observation that **bounded** objects allow some verbs to appear with span modifiers. The verb *eat*, for instance, provides a clear example of resource consumption: it becomes telic in combination with count nouns (*a sandwich, three sandwiches*), quantified mass nouns (*a pound of cheese*) or specific mass nouns (*the cheese*).

When no specific bound is placed on the resource, as with bare plurals (*sandwiches*) and unbounded nouns (*cheese*), the resource cannot be depleted and therefore fails to provide a goal. Note that specific prepositions may force a particular reading: compare *John walked to the park* with *John walked in the park.*

This analysis explains why arguments that do not provide a depletable resource fail to induce atelic readings. The classic example from Vendler is that of *push the cart*, whose object is both bounded and specified, yet fails to serve as a goal, as in *He pushed the cart for an hour* (cf. *He washed the cart in an hour*). Since the PUSH x-schema requires the agent (and energy source) to apply continuous force to the cart, the two participants move together and there is no movement of the agent *with respect to* the cart.

Verbs of creation provide an interesting case of conflicting requirements of a construction and temporal modifier:

3. a. I {read/?wrote} the book for an hour.
   b. I wrote poetry for an hour.

The phrase *for an hour* modifies the ongoing transition without asserting that the finish transition has been taken, but the object of *write* doesn't exist until then, accounting for varying judgments of acceptability. No specific goal is present for the creation of the unbounded noun *poetry* in (b), so durative modification is clearly acceptable.

A similar phenomenon in (4) demonstrates interaction with constructional constraints: although in (4a) *cake* cannot function as a goal, the dative construction in (4b) entails that that the (finished) cake is transferred to the recipient. Again, requirements of durative modification conflict with the entailed creation of the product.

4. a. I baked the cake for an hour.
   b. *I baked you the cake for an hour.

Inter-schema relations

In addition to aspectual composition, the CONTROLLER also brings insight to temporal relations between events. M&S have argued that temporal adverbials such as *when*, *before*, and *after* must be defined in terms of "such notions as causation and consequence, rather than on purely temporal primitives". The fact that *when* does not literally mean "at the same time as" is illustrated by the following example, which M&S take from (Ritchie, 1979):

5. When they built the 39th Street bridge...
   a. a local architect drew up the plans.
   b. they used the best materials.
   c. they solved most of their traffic problems.

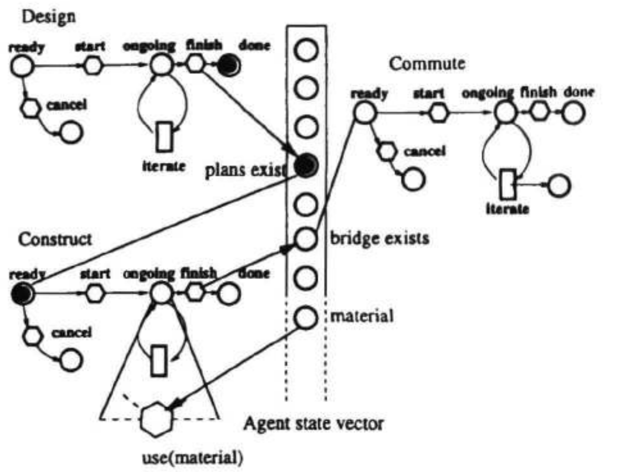

Figure 7: Inter-schema interactions

As this example shows, the event indicated by a *when* clause can in fact occur before, during or after the event indicated by the main clause. M&S propose that the true meaning of *when* is one of a "contingency" relation between the two events. That is, one event depends on the other in a general way.

In our implemented model of x-schema composition, we are able to formally define the notion of contingent relations. The central idea behind our compositional theory is that transitions in the controller graph **set**, **get** and **modify** values of the Agent's state. The controller graph allows us to distinguish between cases of *sequential* and *concurrent* x-schema triggering or inhibition. Additionally, we are able to model the case where the execution of an x-schema is able to *interrupt*, *terminate*, or otherwise modify the execution trajectory of another x-schema.

In Figure 7, we depict the relations between the three x-schemas that correspond to design, construction and use of the bridge in (4). The three x-schemas bind to the specific instance of bridge (corresponding to a specific token). As depicted in Figure 7, the completion of the design process results in the presence of plans, which is a necessary precondition for the construction of the bridge. Similarly, the completion of the bridge is a precondition for its use by commuters. In addition, in our model the presence of appropriate materials is a resource input that gets consumed in the bridge-building process as part of a sub-schema. The interpretation of *before* in the x-schema model is that it indicates a transition (or series of transitions) from the finish node of the action indicated by the subordinate clause to some precondition of the ready node of the main clause. The connection is generally complex, rather than consisting of a single transition.

In the case of *when*, a number of possible enablement relations are possible. Specifically, as illustrated by the example, *when* may have a representation similar to *before*, *after* or *during*. This is not because the word *when* is ambiguous between three meanings, but rather because it simply indicates that there is a causal connection between the two x-schemas, without specifying what type. In our current implementation, the specific interpretation is found by a reachability simulation from the individual nodes of the controller for the action indicated by the main clause to the enabled node of the action indicated by the subordinate clause. These paths indicate whether the two actions may have been simultaneous, or if there was an explicit causal connection.

Acknowledgments

This work was supported in part by NSF grant IRI-9619293 and an NSF Graduate Research Fellowship. Thanks to Jerry Feldman and the NTL group for valuable comments on the paper and its underlying ideas.

References

Bailey, D. R., Feldman, J. A., Narayanan, S., and Lakoff, G. (1997). Modeling embodied lexical development. In *Proceedings of the 19th Cognitive Science Society Conference.*

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge University Press, Cambridge.

Dowty, D. (1979). *Word Meaning and Montague Grammar.* D. Reidel, Dordrecht, W. Germany.

Hwang, C. L. and Schubert, L. K. (1994). Interpreting tense, apsect and time adverbials: A compositional, unified approach. In *First International Conference on Temporal Logic.*

Langacker, R. W. (1991). *Concept, Image, and Symbol.* Mouton de Gruyter, Berlin and New York.

Narayanan, S. (1997a). *Knowledge-based Action Representations for Metaphor and Aspect (KARMA).* PhD thesis, Computer Science Division, EECS Department, University of California at Berkeley.

Narayanan, S. (1997b). Talking the talk is like walking the walk: A computational model of verbal aspect. Proceedings of the 19th Cognitive Science Society Conference.

Reisig, W. (1985). *Petri Nets.* Springer Verlag.

Ritchie, G. (1979). Temporal clauses in English. *Theoretical Linguistics*, 6.

Steedman, M. (1996). Temporality. In *Handbook of Logic and Language.* Elsevier, North Holland.

Verkuyl, H. J. (1993). *A Theory of Aspectuality.* Cambridge University Press.