Representation of Actions as an Interlingua

Karin Kipper and Martha Palmer
University of Pennsylvania
200 South 33rd Street
Philadelphia, PA 19104 USA
{kipper,mpalmer}@linc.cis.upenn.edu

Abstract

We present a *Parameterized Action Representation* (PAR) that provides a conceptual representation of different types of actions used to animate virtual human agents in a simulated 3D environment. These actions involve changes of state, changes of location (kinematic) and exertion of force (dynamic). PARs are hierarchical, parameterized structures that facilitate both visual and verbal expressions. In order to support the animation of the actions, PARs have to make explicit many details that are often underspecified in the language. This detailed level of representation also provides a suitable pivot representation for generation in other natural languages, i.e., a form of interlingua. We show examples of how certain divergences in machine translation can be solved by our approach focusing specifically on how verb-framed and satellite-framed languages can use our representation.

1 Introduction

In this paper, we describe a *Parameterized Action Representation* (PAR) (Badler et al., 1999) that provides a conceptual representation of different types of actions used to animate virtual human agents in a simulated 3D environment. These actions involve changes of state, changes of location (kinematic) and exertion of force (dynamic). PARs are hierarchical, parameterized structures that facilitate both visual and verbal expressions (Badler et al., 2000). In order to support the animation of the actions, PARs have to make explicit many details that are often underspecified in the language. This detailed level of representation is well suited for an interlingua for machine translation applications, since the animations of actions – and therefore the PARs that control them – will be equivalent for the same actions described in different languages. These representations can be incorporated into a system which uses PAR-based animations as a workbench for creating accurate conceptual representations, which can map to seeral different languages as well as produce faithful animations.

The verb classes we are currently considering in this light involve explicit physical actions such as those expressed in the motion verb class and contact verb class (Levin, 1993). Since we are employing PAR as an interlingual representation, we will show examples of how it can handle certain divergences in machine translation, focusing specifically on how *verb-framed* and *satellite-framed* languages (Talmy, 1991) can yield equivalent actions in this representation.

2 PAR representation

We use *parameterized action representations* to animate the actions of virtual human agents. The PAR for an action includes the action's *participants* (its agent and objects), [1] as well as kinematic properties such as its *path*, *manner* and *duration*, and dynamic properties, such as its *speed* and *force* (see Fig. 1). The representation also allows for traditional state-space properties of actions, such as *applicability conditions* and *preparatory actions* that have to be satisfied before the action can be executed, and *termination conditions* and *post assertions* which determine when an action is concluded and what changes it makes to the environment state.

We created a hierarchy of actions, exploiting the idea that verbs can be represented in a lattice that allows semantically similar verbs, such as motion verbs or verbs of contact, to be closely associated with each other under a common parent that captures the properties these verbs all share (Dang et al., 1998). The highest nodes in the hierarchy are occupied by generalized PAR schemas which represent the basic predicate-argument structures for entire groups of subordinate actions. The lower nodes are occupied by progressively more specific schemas that inherit information from the generalized PARs, and can be instantiated with arguments from natural language to represent a specific action such as *John hit the ball with his bat.* The example in Figure 1 is a generalized PAR schema for contact ac-

[1] Objects and agents are stored in a hierarchy and have a number of properties associated with them. Properties of the objects may include their location and status. Agents have capabilities, such as the ability to walk or swim, and properties such as their strength and height.

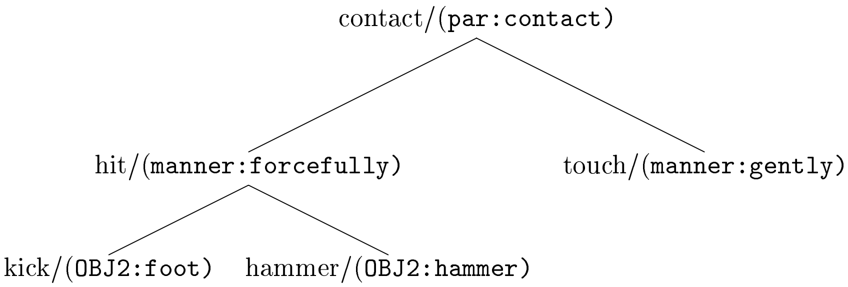

Figure 2: A lexical/semantic hierarchy for actions of contact

$$\begin{array}{c}\text{CONTACT PAR}\\ \left[\begin{array}{l} \textit{activity} : \left[\ \texttt{ACTION}\ \right] \\ \\ \textit{participants} : \left[\begin{array}{lll} \textit{agent}: & \texttt{AGENT} & \\ \textit{objects}: & \texttt{OBJ1}, & \texttt{OBJ2}\end{array}\right] \\ \\ \textit{applic\_cond} : \left[\begin{array}{l}\texttt{reachable(OBJ1)}\\ \texttt{have(AGENT,OBJ2)}\end{array}\right] \\ \\ \textit{preparatory\_spec} : [\texttt{get(AGENT,OBJ2)}] \\ \\ \textit{termination\_cond} : [\texttt{contact(OBJ1,OBJ2)}] \\ \\ \textit{post\_assertions} : [\texttt{contact(OBJ1,OBJ2)}] \\ \\ \textit{path}, \textit{duration}, \textit{motion}, \textit{force} \\ \\ \textit{manner} : \left[\ \texttt{MANNER}\ \right]\end{array}\right]\end{array}$$

Figure 1: A PAR schema for actions of contact

tions between two objects. This schema specifies that the 'contact' action has an agent and two objects, and that the action is concluded when the two objects come together.[2] The preparatory specification of getting the second object is tested and carried out if the object is not possessed. In order to describe a specific action, say *hammer*, we would combine all of its ancestor representations in the action hierarchy, as shown in Figure 2, and add the information specific to that action. Since *hammer* inherits from the PAR for *hit*, and ultimately from the PAR for *contact*, its representation would use the generalized 'contact' PAR, with a forceful manner, and a hammer as the instrument. The action *hit* does not specify any instrument, but inherits the forceful manner and generalized contact PAR from its ancestors, and the action *contact* leaves both the instrument and the manner unspecified, and is associated only with the generalized contact PAR.

The PAR is intended to provide slots for information that is typically conveyed in modifiers or adjuncts in addition to internal verb arguments. As such, it is often the case that several different syntactic realizations can all map to the same PAR schema. For example, *John hit the ball*, *John hit the ball with a bat* and *John swung mightily and his bat hit the ball with a resounding crack* would all map to the same schema.[3]

3 Generating Animations

The main components of our animation system are: a natural language interface, a planner and a graphical animation (see Figure 3). The PARs are used as intermediate representations of the actions between components.

An instruction in natural language starts the process. We use a Synchronous Tree Adjoining Grammar (Shieber and Schabes, 1990; Shieber, 1994) for parsing natural language instructions into derivations containing predicate-argument dependencies (Schuler, 1999). The synchronous parser extracts these predicate-argument structures by first associating each word in an input sentence with one or more *elementary trees*, which are combined into a single derivation tree for the entire input sentence using the constrained operations of *substitution* and *adjunction* in the Tree Adjoining Grammar formalism (Joshi, 1985; Joshi, 1987). As the parser assembles these elementary tree predicates into a predicate-argument structure, it simultaneously selects and assembles the corresponding schemas. It fills in the participants and modifiers, and outputs the PAR schema for the instruction. These schemas may be underspecified for actions such as 'enter' or 'put' and thus not provide enough information for the animation to be produced directly.

[2] In this example, the second object is the instrument with which the action is performed.

[3] The relationship between PARs and alternations may become much more complicated when we consider other verb classes such as change of state verbs.

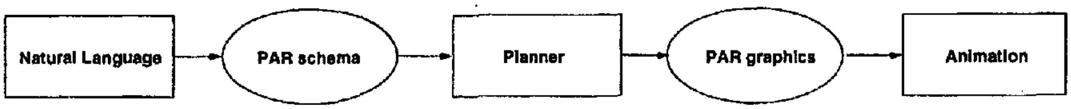

Figure 3: General architecture of the animation system

The planner uses information from the general schema, such as pre-conditions and post-assertions, as well as information derived from the agents' capabilities and the objects properties to fill in these gaps in several ways:

- to select the way (activity) in which the instruction is performed (enter by walking, by swimming, etc.);
- to determine the prepartory actions that must be completed before the instruction is carried out, (for example, in order for an agent to open the door, the door has to be reachable and that may involve a locomotion process);
- to decompose the action into smaller units (put the glass on the table, involves getting the glass, planning a route to the table, etc.)

The output of the planner for the input instruction is a complete description of the actions involved, including participants, preparatory specifications, termination conditions, manner, duration, etc. Participants bring with them a list of inherent properties of the agent (e.g. agent capabilities) or physical objects (e.g., object configurations) and other characteristics, such as 'how to open' for an object such as a door. This complete description refers to a set of animation PARs which can be immediately animated.

In this way, a PAR schema for the action *enter* may actually translate into an animation PAR for *walking into a certain area.* One way to differentiate between action PAR schemas and instantiated animation PARs is to consider what it is possible to motion capture[4] (by attaching sensors to a moving human figure). For example, the *enter* action and the *put* action are quite general and underspecified and could not be motion captured. However, characteristic activities such as *walking* and *swimming* could be. For further details about the animation PARs and the animation system see (Badler et al., 1999) and (Bindiganavale et al., 2000).

4 PAR as an IL

The PAR representation for an action can be seen as a general template. PAR schemas include, as part of the basic sub-categorization frame, properties of the action that can occur linguistically either as the main verb or as adjuncts to the main verb phrase. This captures problems of divergences, such as the ones described by Talmy (Talmy, 1991), for verb-framed versus satellite-framed languages.

New information may come from a sentence in natural language that modifies the action's inherent properties, such as in *John hit the ball slowly,* where 'slowly' is not part of the initial representation of the action 'hit'. This new information is added to the PAR schema.

Verb- versus Satellite-framed languages

Verb-Framed Languages (VFL) map the motion (path or path + ground location) onto the verb, and the manner either onto a satellite or an adjunct, while Satellite-Framed Languages (SFL) map the motion into the satellite, and the manner onto the main verb.

English and other Germanic languages are considered satellite-framed languages, expressing the path in the satellite; Spanish, among other Romance languages, is a verb-framed language and expresses the path in the main verb. The pairs of sentences (1) and (2) from Talmy (1991) show examples of these divergences. In (1), in English, the exit of the bottle is expressed by the preposition *out,* in Spanish the same concept is incorporated in the main verb *salir* (to exit). In (2), the concept of *blowing out* the candle is represented differently in English and Spanish.

(1) *The bottle floated out*
La botella salió flotando
(the bottle exited floating)

(2) *I blew out the candle*
Apagué la vela soplándola
(I extinguish the candle blowing)

4.1 Motion

In order to capture generalizations about motion actions, we have a generalized PAR schema for motion, and our hierarchy includes different types of motion actions such as inherently directed motion and manner of motion actions that inherit from the more general schema, as shown in Figure 4. Directed motion actions, such as enter and exit, don't bring with them the manner by which the action is carried out but they have a inherent termination condition. For example, 'enter a room' may be done by walking, crawling or flying depending on the agents' ca-

[4] There are several other ways to generate motions, for example, through inverse kinematics, dynamics and key-framing.

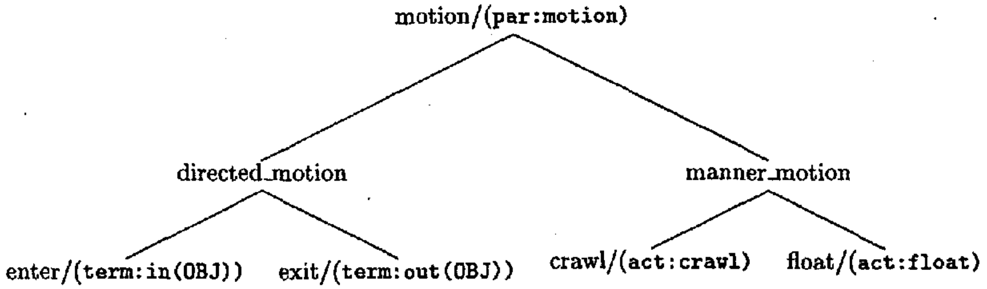

Figure 4: PAR schema hierarchy for motion actions

pabilities, but it should end when the agent is in the room. In contrast, manner of motion verbs express the action explicitly and don't have an intrinsic termination condition.

Motion is a type of framing event where the path is in the main verb for VFLs and in the satellite for SFLs. In (3), we see the English sentence expressing the 'enter' idea in the preposition *into* whereas the Spanish sentence expresses it in the main verb *entrar* (to enter).

(3) *The bottle floated into the cave*
La botella entró flotando a la cueva
(the bottle entered floating the cave)

The PAR schemas don't distinguish the representation for these sentences, because there is a single schema which includes both the manner and the path without specifying how they are realiized linguistically. Mappings from the lexical items to the schemas or to constraints in the schemas can be seen in Figure 5.[5] Independent of which is the source language, the PAR schema selected is `motion`, the *activity* field, which determines how the action is performed (in this case, by floating), is filled by *float* (the main verb in English, or the adjunct in Spanish). The termination condition, which says that action ends when the agent is in the object, is added from the preposition in English and is part of the semantics of the main verb *to enter* in Spanish.

EN float/[par:motion,activity:float]
into/[term:in(AG,OBJ)]

SP entrar/[par:motion,term:in(AG,OBJ)]
flotar/[activity:float]

Figure 5: Entries for the example sentences in (3)

Because all of the necessary elements for a translation are specified in this representation, it is up to the language specific component to transform it into a surface structure that satisfies the grammatical principles of the destination language.

MOTION PAR

$$\left[\begin{array}{l} \textit{activity}: \texttt{float} \\ \\ \textit{participants}: \left[\begin{array}{ll}\textit{agent}: & \texttt{bottle}\\ \textit{object}: & \texttt{cave}\end{array}\right] \\ \\ \textit{termination\_cond}: \texttt{in(bottle, cave)}\end{array}\right]$$

Figure 6: A (simplified) PAR schema for the sentences in (3)

Comparison with other work

Our approach now diverges considerably from the approach outlined in Palmer et al. (1998) which discusses the use of Feature-Based Tree Adjoining Grammars, (Joshi, 1985; Vijay-Shanker and Joshi, 1991) to capture generalizations about manner-of-motion verbs. They do not propose an interlingua but use a transfer-based mechanism expressed in Synchronous Tree Adjoining Grammars to capture divergences of VFL and SFL through the use of semantic features and links between the grammars. The problem of whether or not a prepositional phrase constitutes an argument to a verb or an adjunct (described by Palmer et al.) does not constitute a problem in our representation, since all the information is recovered in the same template for the action to be animated.

The PAR approach is much more similar to the Lexical Conceptual Structures (LCS) approach, (Jackendoff, 1972; Jackendoff, 1990), used as an interlingua representation (Dorr, 1993). Based on the assumption that motion and manner of motion are conflated in a matrix verb like *swim*, the use of LCS allows separation of the concepts of motion, direction, and manner of motion in the sentence *John swam across the lake*. Each one of these concepts is

[5] A lexical item may have several mappings to reflect its semantics. For instance, *float* in English can be used also in the non-motion sense, in which case there will be two entries to capture that distinction.

represented separately in the interlingua representation, as GO, PATH and MANNER, respectively. Our approach allows for a similar representation and the end result is the same, namely that the event of *swimming across the lake* is characterized by separate semantic components, which can be expressed by the main schema and by the *activity* field. In addition, our representation also incorporates details about the action such as applicability conditions, preparatory specifications, termination conditions, and adverbial modifiers. It is not clear to us how the LCS approach could be used to effect the same commonality of representation.

4.2 Instrument

The importance of the additional information such as the termination conditions can be more clearly illustrated with a different set of examples. Another class of actions that presents interesting divergences involves instruments where the instrument is used as the main verb or as an adjunct depending on the language. The sentence pair in (4) shows this divergence for English and Portuguese. Because Portuguese does not have a verb for *to spoon*, it uses a more general verb *colocar* (to put) as the main verb and expresses the instrument in a prepositional phrase. Unlike directed motion actions, a *put with hand-held instrument* action (e.g., spoon, scoop, ladle, etc.) leaves the *activity* field unspecified in both languages. The specific action is generated by taking the instrument into account. A simplified schema is shown in Figure 7.

(4) *Mary spoons chocolate over the ice cream*
Mary coloca chocolate sobre o sorvete com a colher
(Mary puts chocolate over the ice cream with a spoon)

PUT3 PAR

$$\begin{bmatrix} \textit{activity} : - & \\ \textit{participants} : & \begin{bmatrix} \textit{agent} : & \texttt{Mary} \\ \textit{objects} : & \texttt{chocolate,} \\ & \texttt{icecream,} \\ & \texttt{spoon} \end{bmatrix} \\ \textit{preparatory\_spec} : \texttt{get(Mary, spoon)} & \\ \textit{termination\_cond} : \texttt{over(chocolate, icecream)} & \end{bmatrix}$$

Figure 7: Representation of the sentences in (4)

Notice that the only connection between *to spoon* and its Portuguese translation would be the termination condition where the object of the verb, *chocolate*, has a new location which is *over the ice cream.*

5 Conclusion

We have discussed a parameterized representation of actions grounded by the needs of animation of instructions in a simulated environment. In order to support the animation of these instructions, our representation makes explicit many details that are often underspecified in the language, such as start and end states and changes in the environment that happen as a result of the action.

Sometimes the start and end state information provides critical information for accurate translation but it is not always necessary. Machine translation can often simply preserve ambiguities in the translation without resolving them. In our application we cannot afford this luxury. An interesting question to pursue for future work will be whether or not we can determine which PAR slots are not needed for machine translation purposes.

Generalizations based on action classes provide the basis for an interlingua approach that captures the semantics of actions without committing to any language-dependent specification. This framework offers a strong foundation for handling the range of phenomena presented by the machine translation task.

The structure of our PAR schemas incorporate into a single template the kind of divergence presented in verb-framed and satellite-framed languages. Although not shown in this paper, this representation can also capture idioms and non-compositional constructions since the animations of actions – and therefore the PARs that control them – must be equivalent for the same actions described in different languages.

Currently, we are also investigating the possibility of building these action representations from a class-based verb lexicon which has explicit syntactic and semantic information (Kipper et al., 2000).

Acknowledgments

The authors would like to thank the Actionary group, Hoa Trang Dang, and the anonymous reviewers for their valuable comments. This work was partially supported by NSF Grant 9900297.

References

Norman I. Badler, Martha Palmer, and Rama Bindiganavale. 1999. Animation control for real-time virtual humans. *Communications of the ACM*, 42(7):65–73.

Norman I. Badler, Rama Bindiganavale, Jan Allbeck, William Schuler, Liwei Zhao, and Martha Palmer, 2000. *Embodied Conversational Agents*, chapter Parameterized Action Representation for Virtual Human Agents. MIT Press. to appear.

Rama Bindiganavale, William Schuler, Jan M. Allbeck, Norman I. Badler, Aravind K. Joshi, and

Martha Palmer. 2000. Dynamically altering agent behaviors using natural language instructions. *Fourth International Conference on Autonomous Agents*, June.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective levin classes. In *Proceedings of COLING-ACL98*, pages 293–299, Montreal, CA, August.

Bonnie J. Dorr. 1993. *Machine Translation: A View from the Lexicon.* MIT Press, Boston, MA.

R. Jackendoff. 1972. *Semantic Interpretation in Generative Grammar.* MIT Press, Cambridge, Massachusetts.

R. Jackendoff. 1990. *Semantic Structures.* MIT Press, Boston, Mass.

Aravind K. Joshi. 1985. How much context sensitivity is necessary for characterizing structural descriptions: Tree adjoining grammars. In L. Karttunen D. Dowty and A. Zwicky, editors, *Natural language parsing: Psychological, computational and theoretical perspectives*, pages 206–250. Cambridge University Press, Cambridge, U.K.

Aravind K. Joshi. 1987. An introduction to tree adjoining grammars. In A. Manaster-Ramer, editor, *Mathematics of Language.* John Benjamins, Amsterdam.

Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *submitted to AAAI.*

Beth Levin. 1993. *English Verb Classes and Alternation, A Preliminary Investigation.* The University of Chicago Press.

Martha Palmer, Joseph Rosenzweig, and William Schuler. 1998. Capturing Motion Verb Generalizations with Synchronous TAG. In Patrick St. Dizier, editor, *Predicative Forms in NLP.* Kluwer Press.

William Schuler. 1999. Preserving semantic dependencies in synchronous tree adjoining grammar. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL '99).*

Stuart M. Shieber and Yves Schabes. 1990. Synchronous tree adjoining grammars. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING '90)*, Helsinki, Finland, August.

Stuart M. Shieber. 1994. Restricting the weak-generative capability of synchronous tree adjoining grammars. *Computational Intelligence*, 10(4).

Leonard Talmy. 1991. Path to realization–via aspect and result. In *Proceedings of the 17th Annual Meeting of the Berkeley Linguistic Society*, pages 480–519.

K. Vijay-Shanker and Aravind Joshi. 1991. Unification based tree adjoining grammars. In J. Wedekind, editor, *Unification-based Grammars.* MIT Press, Cambridge, Massachusetts.