Action representation for NL instructions

Barbara Di Eugenio*
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA
dieugeni@linc.cis.upenn.edu

1 Introduction

The need to represent actions arises in many different areas of investigation, such as philosophy [5], semantics [10], and planning. In the first two areas, representations are generally developed without any computational concerns. The third area sees action representation mainly as functional to the more general task of reaching a certain goal: actions have often been represented by a predicate with some arguments, such as *move(John, block1, room1, room2)*, augmented with a description of its effects and of what has to be true in the world for the action to be executable [8]. Temporal relations between actions [1], and the *generation* relation [12], [2] have also been explored.

However, if we ever want to be able to give instructions in NL to active agents, such as robots and animated figures, we should start looking at the characteristics of action descriptions in NL, and devising formalisms that should be able to represent these characteristics, at least in principle. NL action descriptions are complex, and so are the inferences the agent interpreting them is expected to draw.

As far as the complexity of action descriptions goes, consider:

Ex. 1 *Using a paint roller or brush, apply paste to the wall, starting at the ceiling line and pasting down a few feet and covering an area a few inches wider than the width of the fabric.*

The basic description *apply paste to the wall* is augmented with the *instrument* to be used and with *direction* and *extent* modifiers. The richness of the possible modifications argues against representing actions as predicates having a fixed number of arguments.

Among the many complex inferences that an agent interpreting instructions is assumed to be able to draw, one type is of particular interest to me, namely, the interaction between the intentional description of an action - which I'll call the *goal* or the *why* - and its executable counterpart - the *how* [1]. Consider:

Ex. 2 *a) Place a plank between two ladders to create a simple scaffold.*
b) Place a plank between two ladders.

In both a) and b), the action to be executed is *"place a plank between two ladders"*. However, Ex. 2.b would be correctly interpreted by placing the plank *anywhere* between the two ladders: this shows that in a) the agent must be inferring the proper position for the plank from the expressed *why "to create a simple scaffold"*.

My concern is with representations that allow specification of both *how's* and *why's*, and with reasoning that allows inferences such as the above to be made. In the rest of the paper, I will argue that a hybrid representation formalism is best suited for the knowledge I need to represent.

2 A hybrid action representation formalism

As I have argued elsewhere based on analysis of naturally occurring data [14], [7], actions - *action types*, to be precise - must be part of the underlying ontology of the representation formalism; partial action descriptions must be taken as basic; not only must the usual participants in an action such as agent or patient be represented, but also means, manner, direction, extent etc.

Given these basic assumptions, it seems that knowledge about actions falls into the following two categories:

1. **Terminological knowledge** about an action-type: its participants and its relation to other action-types that it either specializes or abstracts - e.g. *slice* specializes *cut*, *loosen a screw carefully* specializes *loosen a screw*.

2. **Non-terminological knowledge.** First of all, knowledge about the effects expected to occur

*This research was supported by DARPA grant no. N0014-85 -K0018.

[1] What *executable* means is debatable: see for example [12], p. 63ff.

when an action of a given type is performed. Because effects may occur *during* the performance of an action, the basic aspectual profile of the action-type [11] should also be included. Clearly, this knowledge is not terminological; in

Ex. 3 ***Turn the screw counterclockwise but don't loosen it completely.***

the modifier ***not ... completely*** does not affect the fact that ***don't loosen it completely*** is a ***loosening*** action: only its default culmination condition is affected.

Also, non-terminological knowledge must include information about relations between action-types: temporal, generation, enablement, and testing, where by *testing* I refer to the relation between two actions, one of which is a test on the outcome or execution of the other.

The *generation* relation was introduced by Goldman in [9], and then used in planning by [1], [12], [2]: it is particularly interesting with respect to the representation of *how's* and *why's*, because it appears to be the relation holding between an intentional description of an action and its executable counterpart - see [12].

This knowledge can be seen as *common-sense* planning knowledge, which includes facts such as *to loosen a screw, you have to turn it counterclockwise*, but not *recipes* to achieve a certain goal [2], such as how to assemble a piece of furniture.

The distinction between terminological and non-terminological knowledge was put forward in the past as the basis of hybrid KR system, such as those that stemmed from the KL-ONE formalism, for example KRYPTON [3], KL-TWO [13], and more recently CLASSIC [4]. Such systems provide an assertional part, or A-Box, used to assert facts or beliefs, and a terminological part, or T-Box, that accounts for the meaning of the complex terms used in these assertions.

In the past however, it has been the case that terms defined in the T-box have been taken to correspond to noun phrases in Natural Language, while verbs are mapped onto the predicates used in the assertions stored in the A-box. What I am proposing here is that, to represent action-types, verb phrases too have to map to concepts in the T-Box. I am advocating a 1:1 mapping between verbs and action-type names. This is a reasonable position, given that the entities in the underlying ontology come from NL.

The knowledge I am encoding in the T-box is at the linguistic level: an action description is composed of a verb, i.e. an action-type name, its arguments and possibly, some modifiers. The A-Box contains the non-terminological knowledge delineated above.

I have started using CLASSIC to represent actions: it is clear that I need to tailor it to my needs, because it has limited assertional capacities. I also want to explore the feasibility of adopting techniques similar to those used in CLASP [6] to represent what I called *common-sense* planning knowledge: CLASP builds on top of CLASSIC to represent actions, plans and scenarios. However, in CLASP actions are still traditionally seen as STRIPS-like operators, with pre- and post-conditions: as I hope to have shown, there is much more to action descriptions than that.

References

[1] J. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154, 1984.

[2] C. Balkanski. *Modelling act-type relations in collaborative activity.* Technical Report TR-23-90, Center for Research in Computing Technology, Harvard University, 1990.

[3] R. Brachman, R.Fikes, and H. Levesque. *KRYPTON: A Functional Approach to Knowledge Representation.* Technical Report FLAIR 16, Fairchild Laboratories for Artificial Intelligence, Palo Alto, California, 1983.

[4] R. Brachman, D. McGuinness, P. Patel-Schneider, L. Alperin Resnick, and A. Borgida. Living with CLASSIC: when and how to use a KL-ONE-like language. In J. Sowa, editor, *Principles of Semantic Networks*, Morgan Kaufmann Publishers, Inc., 1990.

[5] D. Davidson. *Essays on Actions and Events.* Oxford University Press, 1982.

[6] P. Devanbu and D. Litman. Plan-Based Terminological Reasoning. 1991. To appear in Proceedings of KR 91, Boston.

[7] B. Di Eugenio. A language for representing action descriptions. Preliminary Thesis Proposal, University of Pennsylvania, 1990. Manuscript.

[8] R. Fikes and N. Nilsson. A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[9] A. Goldman. *A Theory of Human Action.* Princeton University Press, 1970.

[10] R. Jackendoff. *Semantics and Cognition. Current Studies in Linguistics Series*, The MIT Press, 1983.

[11] M. Moens and M. Steedman. Temporal Ontology and Temporal Reference. *Computational Linguistics*, 14(2):15–28, 1988.

[12] M. Pollack. *Inferring domain plans in question-answering.* PhD thesis, University of Pennsylvania, 1986.

[13] M. Vilain. The Restricted Language Architecture of a Hybrid Representation System. In *IJCAI-85*, 1985.

[14] B. Webber and B. Di Eugenio. Free Adjuncts in Natural Language Instructions. In *Proceedings Thirteenth International Conference on Computational Linguistics, COLING 90*, pages 395–400, 1990.