Abstracts

Eren Erdal Aksoy
Inst. f. Physik III, Georg August Univ. Göttingen and BFNT Göttingen

Semantic Analysis of Manipulation Actions: Semantic Event Chains

Defining a generic representation for actions is a hard research topic in cognitive robotics due to high variations in trajectory, pose, and object domains. Conventional methods such as robot programming are far away from coping with all those natural variations in manipulations. Semantic Event Chains (SECs) are introduced as a novel concept to capture only the semantics of manipulation actions independent from all those alterations. SECs essentially extract spatial relations between manipulated objects at time points which are decisive for the manipulation. The SEC representation can further be enriched by incrementally appending observed trajectory, pose, and objects information only at those decisive instants. Hence, the cognitive agent can use the enriched SEC representation to imitate complicated actions even under different circumstances.

Yiannis Aloimonos, K. Pastra and C. Fermüller.
Computer Vision laboratory, University of Maryland, USA

A minimalist grammar for action

Language and action have been found to share a common neural basis and in particular a common ‘syntax’, an analogous hierarchical and compositional organization. While language structure analysis has led to the formulation of different grammatical formalisms and associated discriminative or generative computational models, the structure of action is still elusive. However, structuring action has important implications on action learning and generalization, in both human cognition research and computation. The talk presents a biologically inspired generative grammar of action, which employs the structure-building operations and principles of Chomsky's Minimalist Program as a reference model. In this grammar, action terminals combine hierarchically into temporal sequences of actions of increasing complexity; the actions are bound with the involved tools and affected objects and are governed by certain goals. It is shown how the *tool role* and *the affected-object role* of an entity within an action drives the derivation of the action syntax in this grammar and controls recursion, merge and move, the latter being mechanisms that manifest themselves not only in human language, but in human action too. Examples of parsing are shown in video sequences of manipulation actions, involving objects and tools.

Work supported by the Cognitive systems program of the EU, under the projects POETICON and POETICON++

Tibor Kiss and Antje Müller
Sprachwissenschaftliches Institut, Ruhr-Universität Bochum

Requirements for the annotation of spatial interpretations of prepositions.

Prepositions are generally considered as highly polysemous. This does not only hold for different semantic realms, such as temporality and spatiality, but also within domains. While examples (1, 2) show that the preposition vor can be used to express temporal as well as spatial relations, examples (3, 4, 5) show three interpretations of the preposition über that clearly should be distinguished, and yet can be called spatial. As a consequence, we deem it insufficient to call

prepositions as a whole ‘spatial’ or ‘temporal’.

(1) Er ging vor dem Essen noch mit dem Hund raus.
he left before the lunch once with the dog PART
‘Before lunch, he took the dog for walk one more time.’

(2) Der Teppich liegt vor der Tür.
the carpet lies in--front--of the door
‘The carpet is in front of the door.’

(3) Er hängt das Bild über die Couch.
he hangs the picture above the sofa
‘He hanged the picture above the sofa.’

(4) Er springt über die Mauer.
he jumped across the wall
‘He jumped across the wall.’

(5) Das Bild hängt über dem Wandtresor.
the picture hangs over the strongbox
‘The picture covers the strongbox.’

The annotation schema BSDA (Bochumer Spatial Decision Tree Annotation) is based on decision trees and thus guide the annotator through the application of binary decisions that are mapped to semantic features of an interpretation. It allows the precise annotation of different spatial interpretations of a subset of the German prepositions. It does not only cover different interpretations assigned to the same form, but also relations of identical interpretations assigned to different forms, as is illustrated in (6--9).

(6) Hans und Lisa stehen in der Wiese.
Hans and Lisa stand in the meadow

(7) Hans und Lisa stehen auf der Wiese.
Hans and Lisa stand on the meadow
‘Hans and Lisa are standing on the meadow.’

(8) Hans und Lisa gehen durch die Wiese
 Hans and Lisa walk through the meadow

(9) Hans und Lisa gehen über die Wiese.
Hans and Lisa walk over the meadow
‘Hans and Lisa are walking across the meadow.’

Although the relation between language production and language reception is not directly reflected in an annotation schema, the difference between production and perception must be considered if feature sets for annotations are devised. Most important is to acknowledge that in producing an utterance (as well as in receiving an utterance in a face to face situation), speakers and hearers may employ contextual information and other features that can no longer be detected once the utterance has been turned into a sentence in a corpus.

As a typical example, let us consider so--called reference frames. Levinson (1996) assumes a partition into intrinsic, relative, and absolute reference frames. The distinction between intrinsic and relative reference frames, however, cannot be determined easily if a sentence is encountered in a text. We can illustrate this point with the examples provided by Levinson (1996:137). Levinson classifies (10, 11) as intrinsic, while (12) is classified as relative.

It is striking that it would be almost impossible to tell that (10) in fact uses an intrinsic frame for location, had Levinson not explicitly informed his readers by adding the information in

parentheses. The example in (11) would alternatively allow an interpretation where the ball is located between speaker and hearer, while the hearer shows its back both to the speaker and the ball. One could still utter felicitously The ball is in front of you.

(10) The ball is in front of the chair. (at the chair’s front)

(11) The ball is in front of you.

(12) The ball is in front of the tree.

These examples show that a full set of features for the annotation of spatial senses, that might be available in face--to--face constellations, cannot normally be presumed in the annotation of preposition senses. The set of features at disposal while annotation can thus be described as subset of the full set of features present in human minds during communication; in this sense, a specific annotation of a spatial sense of a preposition may subsume a variety of actual spatial constellations.

This does not only hold for reference frames. Garrod et al. (1999) point out that functional relations may play a role in determining which preposition will be used in an actual constellation. Functional relations are neither independent from object properties (the element to which the arguments of the preposition refer), nor can they be described by ignoring geometrical aspects. Garrod et al. (1999) observe for English on that the relevance of functional aspects decreases the closer the geometrical relation to be described comes to a prototype. Annotators will presumably neither know the geometrical nor the functional setting of the constellation, in which the preposition is used, and they cannot always be inferred from the given context. Hence functional properties cannot be annotated, although they play a role. Again, we see a relation of subsumption between the sense that has been annotated and functional extensions of a sense.

The direct use of BSDA is the annotation of preposition senses, but the annotation of preposition senses is only an intermediary task. The annotations, which form a gold standard, can be used either to classify preposition senses from contexts, or as a grammar for preposition forms, i.e. a mapping from senses to forms. It is in the latter application that it must be kept in mind that annotations can only be provided up to a certain level, and that senses may subsume a variety of appropriate situations to express a certain form.

References:
Garrod, S., Ferrier, G., & Campbell, S. (1999). In and On: investigating the functional geometry of spatial prepositions. Cognition 72, S. 167--189.
Levinson, S. (1996). Frames of reference and Molyneux's question: Cross--linguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Hrsg.), Language and space (S. 109--169). Cambridge: MIT Press.

Norbert Krüger
Sud Dansk University, Mærsk Mc-Kinney Møller Instituttet, Odense, Denmark.

From early vision to symbols

The talk discusses the problem of how to bridge from low-level sensory data to symbolic representations. It is divided into four parts: First, I discuss a definition of a symbol based on two properties: Symbols are condensed and discrete semantic representatives for certain pieces of knowledge (*Expression*) on which operations can be performed expressing relevant functional relations (*Syntax*). In the second part, I will then give an overview about today’s knowledge on the human’s visual system (primarily based on neurophysiological research as outlined in (Kruger et al. (accepted), IEEE PAMI)). I will in particular address the issue of deep hierarchical representations in the visual cortex bridging from low-level visual information to object concept and action. Thirdly, I discuss a suggestion how to define a process in which symbols may emerge

by means of an unsupervised learning scheme. In this context, we suggested in (Koenig and Krueger (2006), Biol. Cybernetics) to combine the three criteria sparseness, predictability and de-correlation in one objective function. I conclude with some remarks on the problem of learning associations between low-level motor-sensory information and symbolic representations.

Stefan Müller
Institut für Deutsche und Niederländische Philologie, Deutsche Grammatik, FU Berlin

Deep Linguistic Analysis, Interfaces and World Knowledge

I will provide the basic building blocks of a linguistic theory and show how constraints from phonology, morphology, syntax, and semantics can be represented. I will comment on the connection to world knowledge, disambiguation via selectional restrictions and non-linguistic information.

Manfred Pinkal
Saarland University, Department of Computational Linguistics

Making TACoS: Grounding Distributional Models of Action Descriptions in Videos

I will present the recently created and newly released Saarbrücken Corpus of Textually Annotated Cooking Scenes (TACoS). The corpus aligns high quality videos with multiple natural-language descriptions of the actions portrayed in the videos. I will report experimental results which demonstrate that a text-based model of similarity between actions improves substantially when combined with visual information. I will also touch upon ongoing research investigating the generation of natural-language descriptions from videos, and offer some speculations about how multimodal corpora might be used to achieve a more fine-grained semantics of action verbs. The work presented in this talk results from an ongoing collaboration with the Visual Processing group of the Max-Planck-Institute for Computer Science.

Reference: Regneri, M., Rohrbach, M., Wetzel, D., Thater, S., Schiele, B., Pinkal, M. (2013): Grounding Action Descriptions in Videos. Transactions of ACL 1 (http://www.transacl.org/wp-content/uploads/2013/03/paper25.pdf)

Daiva Vitkute-Adzgauskiene, Irena Markievicz
Vytautas Magnus University, Dept. Of Computer Linguistics, Kaunas, Lithuania

Unsupervised and semi-supervised learning of action ontology using domain-specific corpora

Task-oriented applications, including robotics applications, often use some ontological framework, representing application domain and acting as a natural language interface. Manual procedures are usually employed for building such ontologies, this resulting in an expensive, time-consuming process.
This presentation shows investigation results for unsupervised and semi-supervised extraction of action-based ontologies using domain specific corpus texts for chemistry operations. Preprocessing of corpus texts used for this purpose includes tokenization, POS tagging and shallow parsing. Investigation results point to the fact, that such method can be useful for extracting verb-concepts, filling in instance information for different action objects and defining different semantic associations. However, hierarchical relationships normally cannot be built using only unsupervised corpus-based methods. Semi-supervised approaches, leveraging existing ontologies and knowledge bases will also be discussed.

Stefan Wermter
Dept of Computer Science, University of Hamburg

Integrated Neural Symbolic Knowledge Technologies for Action Semantics

There has been substantial interest and progress in intelligent systems and knowledge technologies in recent years based on new biomimetic processing principles for integrated knowledge-based systems. While in the past robots were successful in traditional industrial environments, new generations of hybrid intelligent agents and robotic systems are now being developed which focus on bio-inspired and cognitive capabilities, including reasoning, learning and language communication. In this talk we study the potential of nature-inspired, in particular hybrid neural and symbolic representations, in order to build new adaptive action systems, learning ambient intelligence systems, multimodal neural agents, and human robot interaction systems. We will give an overview of learning neural symbolic knowledge technologies and robots from a perspective of integrative hybrid intelligent systems and we illustrate some new developments under development in the Knowledge Technology lab (http://www.informatik.uni-hamburg.de/WTM/).