Mechanisms Meet Content: Integrating Cognitive Architectures And Ontologies

Alessandro Oltramari, Christian Lebiere
Department of Psychology, Carnegie Mellon University
{aoltrama, cl}@andrew.cmu.edu

Abstract

Historically, approaches to human-level intelligence have divided between those emphasizing the mechanisms involved, such as cognitive architectures, and those focusing on the knowledge content, such as ontologies. In this paper we argue that in order to build cognitive systems capable of human-level event-recognition, a comprehensive infrastructure of perceptual and cognitive mechanisms coupled with high-level knowledge representations is required. In particular, our contribution focuses on an integrated modeling framework (the "Cognitive Engine"), where the learning and knowledge retrieval mechanisms of the ACT-R cognitive architecture are combined with integrated semantic resources for the purpose of event interpretation.

1 Introduction

In general, humans can discriminate physical entities by their topological and morphological features, i.e. position, orientation, shape, configuration of proper parts, as well as at a more complex level, that is in terms of categories (person, nail, hammer), thematic roles (agent, patient, instrument), gestalt schemas (organized perceptual units that are not reducible to properties of their parts[1]), etc. Accordingly, "visual intelligence" can be conceived as the human capability to identify events by means of the relations between the entities in a scene, both at the perceptual and conceptual level. In other words, human interpretation of events emerges as the result of intertwined perceptual, cognitive and inferential processes. Reproducing this capability at the machine level requires a comprehensive infrastructure where low-level perceptual and high-level cognitive processes couple with knowledge representations. In this paper we focus on an integrated model – which we refer to as the "Cognitive Engine" – where the learning and knowledge retrieval mechanisms of the ACT-R cognitive architecture are combined with a semantic resource. We aim at integrating cognitive mechanisms and semantic contents to cope with the dynamics of knowledge flow in event-recognition tasks. This research framework instantiates the general idea that cognitive systems benefit from both the mechanism-centered and knowledge-centered approaches to computationally achieve human level intelligence.

The first approach, historically, has focused on general problem-solving programs (Newell, Shaw, Simon 1959) or architectures, i.e. (Anderson 1983), (Newell, 1990). The second approach, partly arising from the limitations of the first, emphasized the knowledge of the system, especially common-sense knowledge, as the source of intelligence (Douglas et al. 1985). Those paradigms have encountered substantial successes in their own rights, but have up to now not achieved the ultimate goal of human-level intelligence. Moreover, both approaches have largely downplayed the other: systems that focus on mechanisms tend to treat knowledge as something to be engineered in ad hoc, task-specific ways, while those that focus on knowledge rely on narrowly tailored mechanisms to access and leverage their content, often raising unsustainable computational requirements in the process.

In this paper, we argue that those approaches are complementary, and that both of their central aspects, mechanisms and knowledge, need to be addressed systematically in a comprehensive approach to artificial intelligence. Moreover, those two components strongly constrain each other, with learning mechanisms determining which knowledge can be acquired and in which form, and specific knowledge contents providing stringent requirements for mechanisms to be able to access them effectively (Anderson and Lebiere 2003).

In the rest of this paper, we introduce each approach, outline our proposed framework that combines them, and

[1] "A complex perception cannot be explained by the linear sum of the sensations that its parts arouse" (Miller and Buckhout 1973, p.118).

then discuss an on-going work where this hybrid approach is being applied to event-recognition.

2 Mechanisms: cognitive architectures as modules of knowledge production

Cognitive architectures attempt to capture at the computational level the invariant mechanisms of human cognition, including those underlying the functions of control, learning, memory, adaptivity, perception and action. We will focus here on one particular cognitive architecture: ACT-R (Anderson and Lebiere 1998). ACT-R is a modular system: its components include perceptual, motor and declarative memory modules, synchronized by a procedural module through limited capacity buffers.

ACT-R has accounted for a broad range of cognitive activities at a high level of fidelity, reproducing aspects of human data such as learning, errors, latencies, eye movements and patterns of brain activity. Declarative memory (DM) plays an important role in the ACT-R cognitive architecture. At the symbolic level, ACT-R models perform two major operations on DM: 1) accumulating knowledge chunks learned from internal operations or from interaction with the environment and 2) retrieving chunks that provide needed information[2]. The ACT-R theory distinguishes 'declarative knowledge' from 'procedural knowledge', the latter being conceived as a set of procedures (production rules) which coordinate information processing between its various modules[3]: according to this framework, agents accomplish their goals on the basis of declarative representations elaborated through procedural steps (in the form of *if-then* clauses). This distinction between declarative and procedural knowledge is grounded in several experimental results in cognitive psychology regarding knowledge dissociation; major studies in cognitive neuroscience implicate a specific role of the hippocampus in "forming permanent declarative memories" and the basal ganglia in production processes (see Anderson 2007, pp. 96-99, for a general mapping of ACT-R modules and buffers to brain areas and Stocco, Lebiere, Anderson 2010 for a detailed neural model of the basal ganglia's role in controlling information flow between cortical regions).

[2] Both chunk learning and retrieval are performed through limited capacity buffers that constrain the size and capacity of the chunks in DM.

[3] In the ACT-R theory, these procedures based on condition-action structures are considered as units for skill acquisition (Anderson and Lebiere 1998, p. 26).

3 Contents: semantic specifications of declarative knowledge

The separation between declarative and procedural knowledge has been an important issue for AI over the years. In 1980 John McCarthy first realized that, in order to enable full-fledged reasoning capabilities, logic-based intelligent systems need to incorporate "re-usable declarative representations that correspond to objects and processes of the world" (McCarthy 1980). Along these lines, John Sowa acknowledged the relevant role played by philosophy in defining a structured representation of world entities (Sowa 1984), i.e. an 'ontology'. According to Guarino, "an ontology" is a language-dependent cognitive artifact, committed to a certain conceptualization of the world by means of a given language (Guarino 1998). Besides the *protocol layer*, where the syntax of the communication language is specified, the ontological layer contains the *semantics* of that language: if concepts are described in terms of lexical semantics, ontologies take the simple form of *dictionaries* or *thesauri*; when ontological categories and relations are expressed in terms of axioms in a logical language, we talk about *formal ontologies*; if logical constraints are then encoded in a computational language, formal ontologies turn to *computational ontologies*[4]. In the framework of cognitive architectures, ontologies play the role of "semantic specifications of declarative knowledge". In this paper we aim at extending ACT-R with a scalable, reusable knowledge model that can be applied across a wide range of tasks. Considering the state of the art[5], most research efforts have focused on designing methods for mapping large knowledge bases to the ACT-R declarative module. Here we commit on taking an integrated approach: instead of tying to a single ontology, we propose to build a *hybrid computational ontology*[6] that combines different semantic dimensions of declarative representations. Our project consists in linking distinctive lexical databases like WordNet (Fellbaum 1998) and FrameNet (Ruppenhofer et al. 2006) with a suitable computational ontology of events, tying the resulting semantic resource to ACT-R cognitive mechanisms. In particular, the following sections describe the fundamental features of an integrated cognitive model for high-level

[4] E.g., Ontology Web Language (OWL). OWL is based on description logics; description logics are *decidable* fragment of First-Order Logic (http://www.w3.org/TR/owl-features/).

[5] For ACT-R, see (Ball, Rodgers, Gluck 2004), (Douglas, Ball, Rodger 2009), (Best, Gerhart, Lebiere 2010), (Emond 2006).

[6] The adjective "hybrid" is used to emphasize the heterogeneity of theories and resources we are adopting for the purposes of the project. For a general survey on hybrid semantic approaches see (Huang et al 2010). For the sake of readability we will henceforth omit the mid-adjective "computational".

visual recognition of events to support visual machine learning with general symbolic representations.

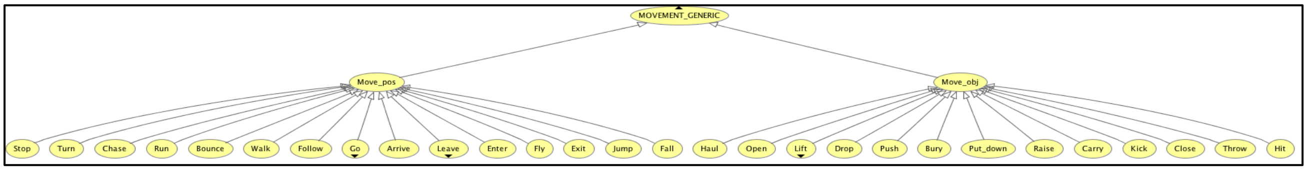

Figure 1 HoMinE's backbone taxonomy of fundamental motor actions

4 Building the *Cognitive Engine*

In the context of the Mind's Eye program[7], we address the perspective of an integrated cognitive model oriented to visual intelligence, outlining methodological aspects and backbone architectural structure. We show how the modular dynamic structures of ACT-R can benefit from augmenting declarative memory with a multi-layered semantic resource, where lexical and ontological knowledge are properly encoded.

4.1 Mind's Eye program: overview

As reported in the Broad Agency Announcement[8], "the Mind's Eye program seeks to develop in machines the capability for visual intelligence: the capability to learn generally applicable and generative representations of action between objects in a scene directly from visual inputs, and then reason over those learned representations". Similarly to what we pointed out in the introduction, this excerpt maintains that event-recognition - far from being a mere sum of observed objects - emerges from a variety of entity-binding relations. A key distinction between this research and the state of the art in computer vision is that the latter exclusively deals with detection and classification of a wide range of objects and their properties (nouns in the description of a scene) while research in visual intelligence should operate at a complementary and more composite level of description: the goal is to look for the perceptual and cognitive underpinnings for recognizing and reasoning about the verbs in a scene, *de facto* analyzing the dynamics of action/object coupling. In the framework of the Mind's Eye program, visual intelligent systems should then be able to reconstruct a "story" from basic constituting events and entities, combining relevant information conveyed by those multiple components into a broader unifying conceptual pattern.

Focusing on architecture development and technological resources, visual intelligent systems generally include three different strata of input/output elaboration: basic optical features (low-level), object detection (mid-level) and event classification (high-level). Notwithstanding the essential "filtering" role played by low-level and mid-level processing, in the remaining sections of the article we just deal with high-level mechanisms and contents, namely the authentic *addenda* of visual intelligence as opposed to pure machine vision. In particular, sections 4.2 and 4.3 outlines the current stage of building the "Cognitive Engine" integrated framework.

4.2 The role of HOMinE

A multi-level conceptual representation of events is needed to elicit and formalize their semantics: understanding the internal structure of event-concepts is a *conditio-sine-qua-non* for enabling mechanism of event-recognition (see next section). As stated in section 3, ontologies can specify meaning at different levels, depending on the focus of representation (language, hierarchical organization, logics, etc.). A full-fledged ontological model, in this sense, should include a comprehensive set of those semantic features, following the line of research of (Oltramari 2001 et al.) and (Gangemi et al. 2003). Grounded on these design principles, HOMinE (Hybrid Ontology for Mind's Eye project) exploits DOLCE (Borgo and Masolo 2009) and OWL-time ontological distinctions as a logical framework of integration for WordNet (WN) and FrameNet (FN).

DOLCE provides the axiomatic basis for developing the formal characterization of action-types and porting it in the framework of Ontology Web Language; OWL-time consists of an OWL standard translation of Allen temporal axioms (Allen 1983), which are considered as state of the art in temporal reasoning. The customization and modularization of these two semantic components allows for a straightforward modeling of core ontological properties of events, such as: i. participation of actors and objects in actions, ii. temporal features of events based on the notions of "instant" and "interval"; iii. temporal measures associated with action instances. In order to enable a suitable temporal reasoning system, we are also

[7] http://www.darpa.mil/i2o/programs/me/me.asp
[8] https://www.fbo.gov/utils/view?id=4ec36649a81c5bcbbd313ba8ae36ca2c (26 March 2010).

augmenting OWL expressivity with SWRL[9] primitives, which can be used to compute arithmetic operations over temporal intervals.

WN is a semantic network whose nodes and arcs are, respectively, *synsets* ("sets of synonym terms")[10] and semantic relations[11]. Over the years, there has been an incremental growth of the lexicon (the latest version, WordNet 3.0, contains about 117K synsets), and substantial enhancements of the entire architecture, aimed at facilitating computational tractability (accordingly, some OWL conversions have been implemented[12]). HOMinE's conceptual layer is based on a partition of WN related to verbs of action, such as "walk", "pick-up", "haul", "kick" chase", etc. In order to find the targeted group of relevant synsets, we basically started from two pertinent top nodes[13], *move#1*[14] and *move#2*[15]. As one can easily notice, the former synset denotes a change of position accomplished by an agent or by an object (with a sufficient level of autonomy), while the latter is about causing someone or something to move (both literally and figuratively). After extracting the sub-hierarchy of synsets related to these generic verbs of action, we have introduced a top-most category "movement-generic", abstracting from the two senses of "move" (see Figure 1). These operations have been performed on Protégé, the most widely used platform for creating computational ontologies in the context of semantic technologies[16]. In order to extract and modify the designated WN partition we used the OntoLing[17] plug-in, a tool that supports semi-automatic lexical population of ontologies. Insofar as lexical databases are augmented with axioms and property restrictions based on OWL primitives, the resulting *hybrid ontologies* can support automatic reasoning.

FrameNet (FN) is the additional conceptual layer of HOMinE. Besides wordnet-like frameworks, a computational lexicon can be designed from a different perspective, for example focusing on *frames*, to be conceived as orthogonal to domains (Pazienza and Stellato 2006). Based on frame semantics (Fillmore 1968), FN aims at documenting "the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses" through corpus-based annotation. Different frames are evoked by the same word depending on different contexts of use: the notion of "evocation" helps in capturing the multi-dimensional character of knowledge structures underlying verbal forms. For instance, if you point to the **bringing** frame, namely an abstraction of a state of affairs where *sentient agents (e.g., persons) or generic carriers (e.g. ships) bring something somewhere along a given path*, you will find several "lexical units"[18] evoking different roles (or frame elements - FEs): i.e., the noun "truck" instantiates the "carrier" role in the frame **bringing**[19]. In principle, the same Lexical Unit (LU) may "evoke" distinct frames, thus dealing with different roles: "truck", for example, can be also associated to the **vehicle** frame ("the vehicles that human beings use for the purpose of transportation"). FN contains about 12K LUs for 1K frames annotated in 150000 sentences.

WN and FN are based on distinct models, but one can benefit from the other in terms of coverage and type of information conveyed. Accordingly, we have analyzed the "evocation" links between the action verbs we have extracted from WN and the related FN frames: those links can be generated through "FN Data search", an on-line navigation tool used to access and query FN[20]. Our study led to a conceptual enrichment of lexical declarative structures for basic action types: starting from WN synset information, and using FN data, we could identify typical roles and fillers of those verbs. This process of extension becomes crucial if one considers the evident isomorphism holding between the elements of ACT-R chunks, namely slots and associated values and elements of frames, i.e. frame elements (roles) and fillers (LUs).

In parallel, we have started to build an ACT-R model for action recognition, suitably expanding its declarative memory by means of HOMinE's semantic layers: section 4.3 outlines this approach.

4.3 The role of ACT-R

We have built an ACT-R model that identifies actions occurring in a simple scenario. The goal is to recognize the most relevant conceptual structures underlying a detected

[9] The acronyms stands for Semantic Web Rule Language: see http://www.w3.org/Submission/SWRL/

[10] Life_form#1 stands for synset {life_form, organism, being, living_thing}, which is identified in the database with a specific code (in this example, {05217061}). Every synset (node of the network) is associated to a gloss (e.g., "the characteristic bodily form of a mature organism").

[11] Hyponymy (sub-class-of), meronymy (part-of), antonymy (opposite-of), troponymy (like hyponymy, but only for verbs), causation, etc.

[12] E.g., http://www.w3.org/TR/wordnet-rdf/

[13] Aka *Unique Beginners* (Fellbaum 1998).

[14] {01835496} <u>move#1</u>**,** travel#1, go#1, locomote#1 (change location; move, travel, or proceed) *"How fast does your new car go?"; "The soldiers moved towards the city in an attempt to take it before night fell";* - <verbs.motion>.

[15] {01850315} <u>move#2</u>, displace#4 (cause to move or shift into a new position or place, both in a concrete and in an abstract sense) *"Move those boxes into the corner, please"; "The director moved more responsibilities onto his new assistant"* - <verbs.motion>.

[16] http://protege.stanford.edu/

[17] http://ai-nlp.info.uniroma2.it/software/OntoLing/

[18] Generically abbreviated with LUs - they correspond to terms in WN synsets.

[19] The sentence is "The *truck bringing* coal to crushing facility at western surface coal mine".

[20] See http://framenet.icsi.berkeley.edu/index.php

action. We hypothesize that 1) *input* are in the form of quasi-propositional descriptions (fed by visual classifiers – a combination of technologies adopted in the *low-level* and *mid-level* strata - that parse the scene and return basic structural descriptions[21]); 2) *output* coincides with the recognition of the correct action pattern and of the related *frames*. The declarative memory of the ACT-R model has been populated with HOMinE, enabling a first level of integration between the semantic layer and the cognitive architecture. Mapping HOMinE to ACT-R requires some preliminary analysis of the basic structures involved. Chunks, typed structures composed of specific slots, are the building blocks of ACT-R declarative memory, while ontologies are based on so-called "categories" ("*object*", "*event*", "*attribute*", "*value*", etc.) and "relations" ("*participation*", "*part-of*", "*dependence*", etc.). Let's consider the following chunk types and chunk instances:

(chunk-type car color) (c1 ISA car color red[22])
(chunk-type race duration) (r1 ISA race duration 1hour)

One can think of ontological categories as mapping to different elements of chunks: objects/events mapping to chunk types (e.g., *car/race*), attributes to slots of chunks (e.g., *color/duration*), and values to fillers of slots (e.g., *red/1hour*). Relations (e.g., *has_color/has_duration*) remain implicit, although they essentially "glue" together those pieces of declarative knowledge (e.g., *car – has_color – red*; *race – has_duration – 1hour*). Alternatively, we can observe that ontological relations can be represented as chunk types as well: e.g., we could have defined *has_duration* as a chunk type with slots *event* and *duration*, with *race* and *time* as filler:

(chunk-type has_duration event duration)
(r1 ISA has_duration event race duration 1hour)

The category *race* would then become filler of the slot *event*. This potentially variable matching between ontological knowledge and declarative representations reflect the fact that chunks are originally seen as units of memories, without any strong ontological constraint: in fact, anything that is introduced in declarative memory is a chunk, no matter whether an object, an event, an attribute, a value or a relation. Chunks are goal-driven, namely they represent the knowledge a person is expected to manipulate to solve a problem, and environment-driven, namely they represent combinations of items occurring together. Thus chunk types should reflect both the environment in which cognition operates and the goals that it pursues. The shift from chunk type to filler addresses the potential of alternative representations of categories in ACT-R. Conversely, from the viewpoint of *hybrid ontologies*, representing relations as chunk types becomes an important requirement: relations enable OWL-based inference engines[23] and definitely demand for an explicit counterpart in the declarative memory of the cognitive agent to make the integration effective. The ACT-R architecture also supports "inheritance"[24] from a single chunk type ("single inheritance"), so that different levels of specialization for slot and values are supplied. "Single inheritance" is a central feature for automatic reasoning over ontologies, since it helps prevent logical inconsistencies and internal incoherence of models (which are typically correlated to "multiple inheritance"). HOMinE discards "multiple inheritance" too, maintaining full compatibility with the ACT-R architectural choice.

In order to elicit semantic and cognitive structures of actions, the general class "micro-action"[25] has been fleshed out and suitably modeled in HOMinE. Micro-actions can be conceived as adequately fine-grained temporal parts of actions. In this sense, micro-actions play the role of atomic components of action: adequate temporal sequences of micro-actions can be associated to so-called "ontological patterns" (Gangemi and Presutti 2009), to be seen as complex schemas of actions. In this context, the notion of "ontological pattern" is complementary to the notion of "frame": a pattern can be seen as a super-structure emerging from the latter. In particular, each micro-action type (extend-arm, grasp object, bend-forward, etc.) evokes a specific "frame" and helps in identifying the roles played by agents and objects in the sub-structural part of the overall scene; when sequences of micro-actions are presented, the ACT-R model disambiguates and interprets their semantic content by means of WordNet and FrameNet structures, finally matching the overall sequence to the most suitable ontological patterns[26]. The actual output of the model is a time course of activation of the various ontological patterns as the model is presented with a sequence of temporal frames (Figure 2). Partial matching based on similarity measures and spreading of activation based on structural relations are the main mechanisms used by the ACT-R model to perform efficient action pattern recognition (Equation 1).

[21] E.g., "Person1-kicks-ball5".
[22] A specific red nuance, not to be confused with the abstract property "redness", which is a sub-type of "color".

[23] Ontological relations correspond to OWL object-properties and data-type properties.
[24] The notion of inheritance corresponds to "IS-A" in Computer Science and "hyponymy" in (computational) lexical semantics.
[25] Alternate notation: μ-action.
[26] WordNet and FrameNet structures and ontological patterns are embedded in the model as declarative chunks.

$$A_i = \ln \sum_j t_j^{-d} + \sum_k W_k S_{ki} + \sum_l MP_l Sim_{li} + N(0,\sigma)$$

Equation 1 Bayesian Activation Pattern Matching

Base-level activations of verbs actions have been derived by frequency analysis of the ANC (American National Corpus)[27] to account for base-rate effects in verb preferences. In addition, we constrain semantic similarity within verbs to "gloss-vector" measure computed over WordNet computational lexicon (Pedersen, Siddharth, Michelizzi 2004). Finally, strengths of associations are set (or learned) by the architecture to reflect the number of patterns to which each micro-action is associated, the so-called "fan effect" controlling information retrieval in many real-world domains (Anderson & Schooler, 1990).

For reasons of space, we limit our presentation to an overview of the model. Let's consider four sample sentences consecutively presented to the ACT-R cognitive model augmented with HOMinE:

(1) Person1 **enters** car1
(2) Person2 **throws** ball1
(3) Car1 **exits** left
(4) Person2 **runs** left

Following the typical schema for sentence processing and representation in ACT-R (Anderson 1974), our model parses the screen, reads sentences and encodes related chunks accordingly. Afterwards, the actual retrieval of HOMinE declarative representations starts: the model retrieves the frames and frame-elements evoked by the quasi-propositional inputs, implicitly assuming that the sequential presentation corresponds to the temporal order underlying the denoted micro-actions. The "retrieve-frame-and-elements" production then encodes the process of frame evocation. The rationale is to search for distinctive instances of frames and frame elements in sentences to unfold their semantic features. In our example, declarative chunks encode conceptual information about the evocation of frames and frame elements:

(e1 ISA evocation verb **enter** frame arriving
fe1 *theme* fe2 *goal*)
(e2 ISA evocation verb **throw** frame cause-motion
fe1 *agent* fe2 *theme*
(e3 ISA evocation verb **exit** frame departing
fe1 *theme* fe2 *direction*)
(e4 ISA evocation verb **run** frame self-motion
fe1 *self-mover* f2 *direction*)

[27] http://americannationalcorpus.org/

When the production fires, the model retrieves the conceptual roles (in *italics*) played by the entities in the scene, for each micro-action. Afterwards, in order to prompt a choice within the available ontological patterns of action, spreading activation can be exploited through the ACT-R sub-symbolic computations (Anderson and Lebiere 1998). Spreading of activation from the contents of verb slots triggers the evocation of chunks related to pattern components in the context of the perceived scene. In this example, three patterns are activated at different times:

(o1 ISA ontology-pattern pattern **enter**
component1 enter)
(o3 ISA ontology-pattern pattern **leave**
component1 enter component2 exit)
(o4 ISA ontology-pattern pattern **chase** component1 enter
component2 exit component3 run)

Respectively, o1 is retrieved after the encoding of (1), o3 is retrieved after the encoding of (3), o4 is retrieved after (4) is encoded – when Person2 runs along the same direction of Car1 (left). The three tiny columns in Figure 2 represent this sequential positive activation. Figure 2 also shows that in correspondence of the second micro-action (labeled as "2 μ-action"), there is no active pattern, meaning that (2) doesn't play any major role in the context of the presented sequence. This information clearly matches the intuition that person2 throwing a ball is not essential for understanding that, afterwards, person2 chases person1. Note that by means of the reasoning capabilities of HOMinE, we can also query SWRL rules to get inferential knowledge about the presented inputs, e.g. "if a person enters a vehicle and the vehicle exits from the scene, then also the person exits the scene". Finally, suppose that input (3) is "Car1 flees left". In the common-sense experience, "flee" should even be a stronger clue than "exit" for the detection of a chasing event. Nevertheless, o3 doesn't contain any "flee" verb component and adding another chase-pattern to account for a single compositional variation would simply "nullify" the advantage of having general patterns of action. Using ACT-R, we can overcome this problem and include variations in patterns by setting an high similarity parameter between the two verbs *exit* and *flee*: in other words, if the model perceives *flee*, it will make reference to the *frame(s)* evoked by *exit* through the ACT-R mechanism of "partial matching", which allows the semantics of similarity between chunks to be reflected in the retrieval process (Budiu and Anderson 2004).

5 Conclusion

This paper presented the general framework of integration between the ACT-R cognitive architecture and semantic resources. In particular, we considered the task of high-level visual recognition of actions, outlining how HOMinE ontological features can augment ACT-R declarative representations. Future work will be devoted to enhance both the semantic layer and the cognitive model: the long-term goal is to have a kind of client-server communication between ACT-R and HOMinE, to foster automatic inferences over relevant features of the action structures. Future releases of the system will be also able to deal with more complex action structures, exploiting hierarchical dependencies between ontological patterns; we are also working on suitable algorithms for natural language generation in order to mimic as much as possible the human-level description of the detected actions.

In the context of Mind's Eye program, improving the ACT-R model and the HOMinE ontology is a major effort towards the development of the "Cognitive Engine", an integrated modeling framework for visual intelligent systems.

Acknowledgments. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0061. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

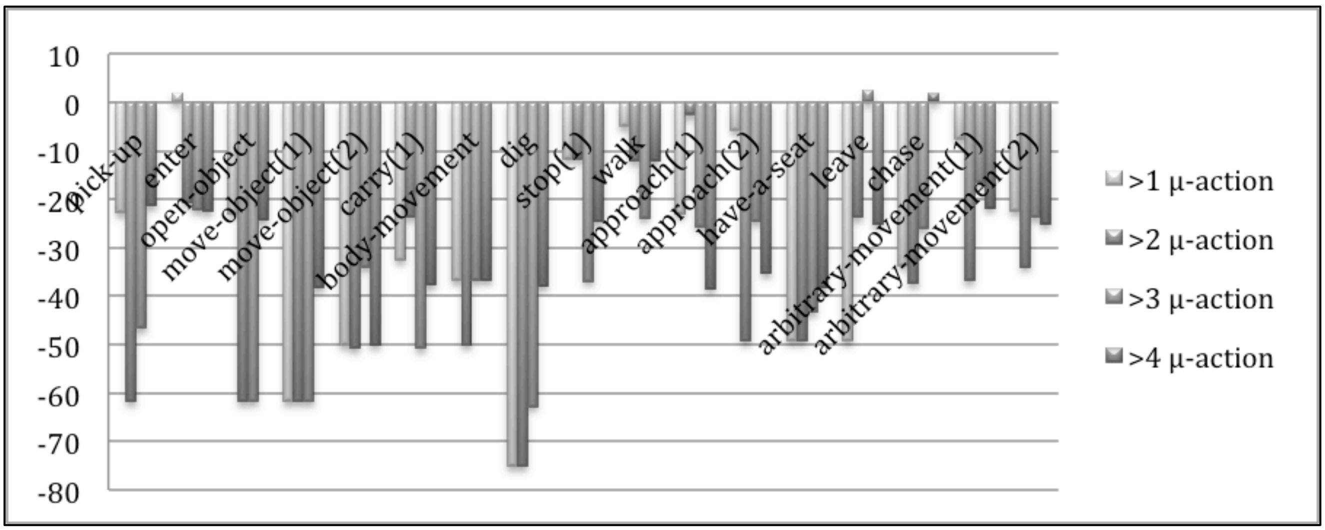

Figure 2 Time course of the model activation for enter, leave and chase ontological patterns

References

Allen J.F. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, Volume 26.

Anderson, J. R. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, J.R. 2007. *How Can the Human Mind Occur in the Physical Universe?* NY: Oxford University Press.

Anderson, J.R. 1974. Retrieval of propositional information from long-term memory. *Cognitive Psychology* 5: 451-474

Anderson, J.R., Lebiere, C. 1998: *The atomic components of thought.* NJ: Erlbaum, Mahwah.

Anderson, J. R., Lebiere, C. L. 2003. The Newell test for a theory of cognition. *Behavioral & Brain Sciences* 26: 587-637.

Schooler, L. , Anderson, J. R. 1990. The disruptive potential of immediate feedback. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 702-708, Cambridge, MA.

Newell, A., Shaw, J.C., Simon, H.A. 1959. Report on a general problem-solving program. In Proceedings of the International Conference on Information Processing, 256-264.

Ball, J., Rodgers, S., Gluck, K. 2004. Integrating ACT-R and Cyc in a large-scale model of language comprehension for use in intelligent agents. In *Papers from the AAAI Workshop*, 19-25,, Menlo Park, Calif.: AAAI Press.

Borgo, S., Masolo, C. 2009. Ontological Foundations of DOLCE In Staab, S., Studer, R. eds., *Handbook on Ontologies* (Second Edition), 361-382, Springer Verlag.

Budiu, R. & Anderson, J. R. 2004. Interpretation-Based Processing: A Unified Theory of Semantic Sentence Processing. *Cognitive Science* 28: 1-44.

Best, B. J., Gerhart, N., Lebiere, C. 2010. Extracting the Ontological Structure of OpenCyc for Reuse and Portability of Cognitive Models. In Proceedings of the 17th Conference on Behavioral Representation in Modeling and Simulation.

Douglass, S., Ball, J., Rodgers, S. 2009. Large declarative memories in ACT-R. In Proceedings of the 9th International Conference of Cognitive Modeling. Manchester, United Kingdom.

Douglas B. Lenat, Mayank Prakash, Mary Shepherd. 1985. CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks. Artificial Intelligence, 6 (4): 65-85.

Emond, B. 2006. WN-LEXICAL: An ACT-R module built from the WordNet lexical database. In: Proceedings of the Seventh International Conference on Cognitive Modeling. Trieste, Italy, 359-360.

Fellbaum, C. (ed.): WordNet - An Electronic Lexical Database. Cambridge, Massachusetts. MIT Press (1998)

Fillmore, C.J. 1968. The case for case. Bach, E., Harms, T. eds. *Universals in Linguistic Theory*. New York: Rinehart and Wiston.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. 2003. Sweetening WordNet with DOLCE. *AI Magazine* 24 (3): 13-24.

Gangemi, A., Presutti, V. 2009. Ontology Design Patterns. Staab, S., Studer, R. eds., *Handbook on Ontologies* (Second Edition), 221-243, Springer Verlag.

Gruber, T.R. 1993. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5: 199–220.

Guarino, N. 1998. Formal Ontology and Information Systems. In Guarino, N. (ed.) *Formal Ontology in Information Systems*, Trento: IOS Press.

Huang, C.R., Calzolari, N., Gangemi, A, Lenci, A., Oltramari, A., Prevot, L. eds. 2010. *Ontology and the Lexicon*. Cambridge University Press.

McCarthy, J. 1980. Circumscription – A Form of Non-Monotonic Reasoning. *Artificial Intelligence* 5, (13): 27–39

Miller, G. and Buckhout, R. 1973. Psychology: the science of mental life. New York: Harper and Row, 2nd edition.

Newell, A.1990. *Unified Theories of Cognition*. Harvard University Press.

Ovchinnikova, E., Vieu, L., Oltramari, A., Borgo, S., Alexandrov, T. 2010. Data-driven and Ontological Analysis of FrameNet for Natural Language Processing. In Proceedings of 7th International Conference on Language Resources and Evaluation, La Valletta, Malta.

Pazienza, M.T., Stellato, A. 2006 An open and scalable framework for enriching ontologies with natural language content. In Huang, C.R., Oltramari, A., Buitelaar, P., Fellbaum, C., Lenci, A. eds. Proceedings of Ontologies and Lexical Resources – workshop at Linguistic Resources and Evaluation Conference, Genoa, Italy.

Pedersen, T. Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity: measuring the relatedness of concepts. In Demonstration Papers at HLT-NAACL 2004, 38-41. Association for Computational Linguistics, Stroudsburg, PA, USA.

Ruppenhofer, J., Ellsworth, M., Petruck, M.R., Johnson, C.R., and Scheffczyk, J. 2006. FrameNet II: Extended Theory and Practice. International Computer Science Institute.

Sowa, J. F. 1984. *Conceptual Structures. Information Processing in Mind and Machine*. Reading: Addison Wesley.

Stocco, A., Lebiere, C., & Anderson, J.R. 2010. Conditional Routing of Information to the Cortex: A Model of the Basal Ganglia's Role in Cognitive Coordination. *Psychological Review* 117(2): 541-574.