

Scientific report after the STSM “Large Scale Knowledge Bases of integrated Vision and Language Representations”.

Lorenzo Gregori

LABLITA, Humanities Dept., University of Florence.

Beneficiary: Lorenzo Gregori

Host: Cognitive Systems Research Institute, Athens, Greece.

Duration: 01.02.2017 - 30.04.2017

COST Action: IC1307: iV&L Net

Purpose of the STSM

The purpose of the mission was the analysis of action concepts by comparing two resources based on different theoretical framework: PRAXICON¹, a multisensory and multimodal semantic memory developed in CSRI in the framework of European research and development Projects, and IMAGACT², a multilingual ontology of action developed within two Italian research Projects (IMAGACT and MODELACT), that contains a fine-grained categorization of concepts and the visual representation of actions. The two resources are both focused of action representation, but from different points of view: IMAGACT action concepts are linguistically discriminated (according to the semantics of verbs in different languages) and are visually represented by video scenes, while PRAXICON concepts are based on the action motorics and are represented by visual and linguistic entities (image, video, words, phrases).

For this reason the analysis of actions through these different resources is very interesting and lead to an exploration of linking and mapping possibilities, in order to put the basis for the creation of an integrated resource, gathering sensorimotor, semantic, lexical and visual information about actions.

Description of the work carried out during the STSM

During the first part of the STSM I became familiar with the PRAXICON model, through the study of the related interdisciplinary literature and the investigation of computational aspects regarding the representation of entities in graph databases. Within this study I set up a PRAXICON Java module to allow the migration of the PRAXICON MySQL database to a Neo4J graph database.

1 <https://github.com/CSRI/PraxiconDB>

2 <http://www.imagact.it>

Then a comparison to investigate the mapping possibilities between IMAGACT and PRAXICON has been conducted. The differences in theoretical foundations of the two resources led us to discard the hypothesis of a mapping between the action concepts: PRAXICON action definition is based on the physical action motorics, while IMAGACT concepts are linguistically motivated. Conversely the visual representations of actions are independent by any theoretical frame: for this reason the IMAGACT video scenes that are prototypical representations of action concepts, can also be used to represent PRAXICON concepts. Starting from this point I developed two main activities during the STSM:

1. The annotation of IMAGACT video scenes through the PRAXICON model;
2. The development of a Java-based video annotation interface.

Within the set of 1.010 video scenes belonging to the IMAGACT ontology, 300 scenes have been annotated on the basis of PRAXICON. In this frame an IMAGACT video scene, that represents a prototype of a lexicalized concept, can work as a visual representation for one or more PRAXICON concepts: in fact in many cases an action that is perceived as unitary (and categorized with a verb) is composed of a set of “smaller” actions performed in a sequence or simultaneously. It is also possible that the scene does not represent any PRAXICON concept, for example in the cases in which the action is not voluntary (like “to fall”), where the lack of goal does not allow to create a PRAXICON action.

The Java annotation interface (PraxiconVideoAnnotation) is built on top of the PRAXICON DB (<https://github.com/CSRI/PraxiconDB>) by taking into account the requirements that came out from the scene annotation activity. This tool integrates data from PRAXICON and WordNet and allows to load a video playlist, play a video and create the corresponding PRAXICON action concepts that the video represent. The tool is freely available at <https://github.com/lablita/PraxiconVideoAnnotation>

Future works and collaborations

The future work plan foresees a collaboration between the LABLITA Research Unity of Florence and the CSRI of Athens. The very next activities comprise:

- Exploiting PraxiconVideoAnnotation to import the annotation of the 300 scenes in the PRAXICON database;
- Performing the annotation of the remaining IMAGACT scenes;
- Updating PraxiconVideoAnnotation to allow the imports of annotated XML files.

After these works we will be able to perform an extensive data analysis of the merged IMAGACT-PRAXICON knowledge base by combining the linguistic and the motoric points of view on the action concepts.

References

- Moneglia, M., Brown, S., Frontini, F., Gagliardi, G., Khan, F., Monachini, M., & Panunzi, A. (2014). The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the representation of lexical encoding of Action. In *LREC* (pp. 3425-3432).
- Moneglia, M., Gagliardi, G., Panunzi, A., Frontini, F., Russo, I., & Monachini, M. (2012). IMAGACT: deriving an action ontology from spoken corpora. In *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-8)* (pp. 42-47).
- Panunzi, A., De Felice, I., Gregori, L., Jacoviello, S., Monachini, M., Moneglia, M. & Russo, I. (2014). Translating Action Verbs Using a Dictionary of Images: The IMAGACT Ontology. In *XVI EURALEX International Congress: The User in Focus* (pp. 1163-1170).
- Pastra, K. (2008). PRAXICON: the development of a grounding resource. In *Proceedings of the international workshop on human-computer conversation*. Bellagio, Italy.
- Pastra, K. & Aloimonos, Y. (2011). The minimalistic grammar of action. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367(1585). The Royal Society.
- Pastra, K., Dimitrakis, P., Balta, E., & Karakatsiotis, G. (2010). PRAXICON and its language-related modules. In *Proceedings of companion volume of the 6th Hellenic conference on artificial intelligence (SETN)*.