MuVAnBA
Multilingual Video Annotation of Basic Actions
Irene Russo (ILC CNR)

research idea

- words denoting actions are essential for second language learning but they are among the most difficult items to master, both at the lexical and at the construction level;
- traditional learners' materials present semantic distinctions relying on dictionaries' definitions and corpus occurrences, with scarce and unsystematic usages of multimodal items such as pictures and videos;
- videos could be very helpful for a proper understanding of words denoting actions, with the possibility to have the same lexical items in different constructions pointing to a set of videos that share a semantic core.

Grounding language learning with multimodal content is an effective strategy that can be implemented if guided crowdsourcing experiments are created with this objective.
MuVAnBA wants to promote the multilingual crowdsourced annotation for short videos depicting actions through a dedicated annotation platform, asking multiple annotators to describe the same video in their mother language. Automatic action segmentation tools will be used in the pre-processing phase, in order to obtain a big dataset of video snippets of unitary actions.

- Since the same action can be described with different level of details and stylistic choices, the annotation will be organized as a game-with-a-purpose aiming at producing the canonical description, i.e. the one that the average mother speaker of a language will produce;
- all the descriptions will be evaluated by a NLP dedicated module able to compute semantic similarity between sentences, in order to find out the set of canonical captions that are semantically close.

With a user-friendly interface the task will enable the creation of additional annotation layers such as:
- a level containing PoS tagging, useful in particular for under-resourced languages that don't have NLP tools for this;
- a level containing links to the agents/objects mentioned in the description and present in the video, highlighted manually with the use of bounding boxes.

The final result of the project will be a multimodal resource containing action denoting words for multiple languages linked to multiple videos and accessible typing the words or searching for pictures (derived extracting bounding boxes from video frames) of agents/objects that can be the agent or the theme of the actions.
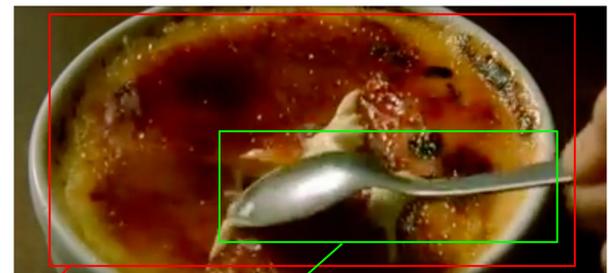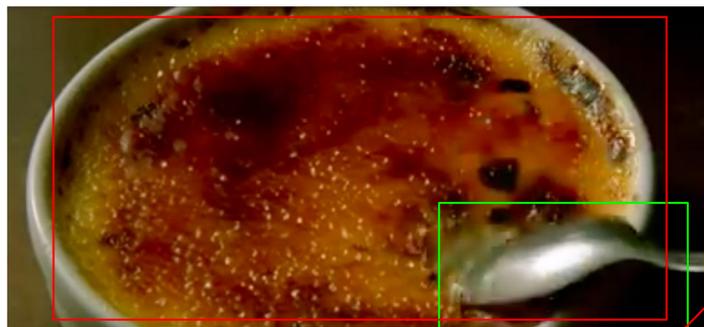
Amelie is breaking the top of the **crème brûlée** with a **spoon.**

my skills

- PhD in computational linguistics
- past experience in annotation experiments (even with crowdsourcing platforms)
- strong belief in the necessity to combine linguistic and multimodal information
- creative and open to new ideas and approaches

(complementary) skills needed

- (a bit of) expertise in annotation and processing of video content
- experience in the creation of crowdsourcing experiments and game-with-a-purpose
- an interest for NLP tools and resources
- positive attitude toward theoretical informed research
- open to new ideas