Action Identification and Local Equivalence of Action Verbs: the Annotation Framework of the IMAGACT Ontology

Massimo Moneglia, Alessandro Panunzi, Lorenzo Gregori
LABLITA, University of Florence
Piazza Savonarola 1 – 50132 – Firenze (Italy)
massimo.moneglia@unifi.it, alessandro.panunzi@unifi.it, lorenzo.gregori@unifi.it

Abstract

Actions are productive concepts, but they are both linguistically and cognitively underdetermined: what defines an action in the event continuum is still an open question. The linguistic encoding of actions offers both problems and solutions to the issue of identifying these concepts. First of all, many action verbs do not identify one single action, but can refer to different action concepts. Secondly, each language categorizes actions in its own way. The IMAGACT Ontology of Action adopts a flexible approach to categorization that allows us to make a semantically coherent discrimination of action concepts across different languages. To this end IMAGACT employed the systematic annotation of Local Equivalence, i.e. the property that different verbs (with different meanings) can refer to the same action concept. However, Local Equivalences alone do not solve the problem of action identification: a further distinction of Local Equivalence relations is required in order to separate productive from non-productive equivalences. In fact, when such productivity is missing, Local Equivalences are not essential for action concept identification.

Keywords: Ontology of Action, Annotation Methodology, Equivalent Verbs

1. The IMAGACT Framework

1.1 Action Concepts and General Verbs

Each action extends to an open set of differing events: therefore, actions are productive concepts. Productivity manifests itself first in the action/object relation. In principle, an action specifies a pattern of world modifications performed by an AGENT. This pattern can be applied to an open set of objects and, conversely, each object may undergo an open set of actions.

Such a property is mirrored at the linguistic level in the predicate/argument structure: a verb referring to an action can be applied to an open set of arguments, which in turn may undergo an open set of action verbs.

Beyond this, action annotation is a complex task, since the reference entities are largely underdetermined: what defines an action in the event continuum remains an open question. For instance, as modern neurology has demonstrated, different sensory-motor patterns performed with the same GOAL are categorized under the same action concept at the brain level (Umilità et al., 2008). This datum would lead us to define the way we conceptualize actions on the basis of AGENT intentions. Nonetheless, we do not really know to what extent we are performing the same action when we use different means to achieve the same purpose. In other words, we do not know to what extent is the GOAL a definitional criterion for a given action concept.

The linguistic encoding of actions offers both problems and solutions to the issue of identifying these concepts. An action verb is usually understood by competent speakers as a tag for one single action, but this impression does not match with reality. A large number of high frequency action verbs, such as *to take* or *to put*, do not identify one single action.

We call verbs which share this property *general*. Such verbs refer to many different action concepts, making the need for a cognitive level of action categorization which is independent from the linguistic one quite clear.

Figure 1, derived from the IMAGACT Cross-Linguistic Ontology of Action (Moneglia, 2014; Panunzi et al., 2014), shows this in practice: the different actions to which we can refer using the verb *to take* are presented by means of screenshots, with each one taken from a brief Scene (i.e. a recorded video or a 3D animation belonging

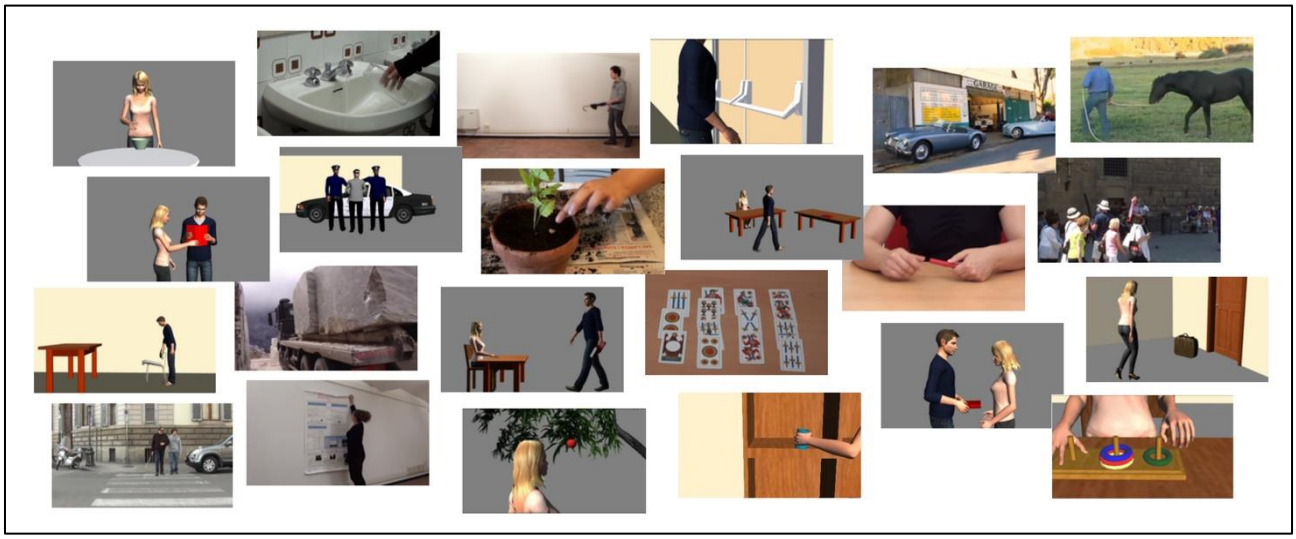

Figure 1: Actions referred by the verb *to take* (partial and unstructured overview)

Figure 2: Differences between *to take* (EN) and *toru* (JP) in the REMOVAL semantic field

to the IMAGACT ontology)[1]. Within the IMAGACT framework, these Scenes are conceived as prototypes (Rosch, 1978; 1983) that stand for broader classes of actions. In this way, the set of Scenes linked to each verb identifies its semantic variation. As Figure 1 shows, the actions of *taking* correspond to many different (cognitively distinguishable) activities within the actual language usage, each one representable by a prototypical Scene.

Contrary to other lexical databases (e.g. WordNet; Miller, 1995; Fellbaum, 1998), IMAGACT records in the ontology only those fields of application in which a verb extends "in its own meaning". Abstract and metaphorical concepts are excluded, even if they are frequently conveyed by action verbs (37,9% of action verb occurrences in the Italian corpus; and 49,9% in the English corpus; Moneglia, 2014b). This requirement ensures that the corpus induced ontology specifically gathers *physical actions* and that verbs apply productively to the action concepts in their extension. This choice is underpinned by a semantic reason: despite the difference among the actions represented in Figure 1, competent speakers can indicate whatever instance of each prototype as "an instance of what we mean by *take*". This cannot be the case with abstract meaning, which undergoes to specific use conditions. For instance, no English speaker will identify the following WordNet synset as a prototypic instance of what we mean by *take*:

> S: (v) assume, acquire, adopt, take on, take (take on a certain form, attribute, or aspect) "His voice took on a sad tone"; "The story took a new turn"; "he adopted an air of superiority"; "She assumed strange manners"; "The gods assume human or animal form in these fables".

In parallel, this concept cannot freely extend to other entities of the same semantic type: even if *she took an air of superiority* works fine, the sentences *she took a bad habit* and *he took the gambling problem* are not acceptable.

Moreover, if we put the question regarding the referential variation of a verbal entry to the cross-linguistic level, we can easily see that each language parses the continuum of action in its own way (Majid et al., 2008; Kopecka and Narasimhan, 2012).

For instance, the Japanese verb *toru* (取る), which roughly corresponds to the concept of *taking*, shows productive differences when compared with the variation of *to take*. In brief, *toru* is not applicable to the action of *bringing something or someone to somewhere* (for which Japanese uses the verb *yoseru*, 寄せる) nor to the simple action of *grasping* (*tsukamu*, 掴む). Conversely, *to take* is not applicable when catching something, which is a frequent use of *toru* (e.g. *Mami ga boru wo toru* 真美がボールを取る; En. *Mary catches the ball*). Moreover, *toru* can be applied to a larger set of events in which something is *removed* (see the examples in Figure 2).

To sum up: action concepts are not determined neither in language nor in cognition in general; action verbs correspond to linguistic concepts, able to refer to more than one cognitive entity; each language categorizes actions in its own way.

In order to manage this complexity, IMAGACT has adopted a flexible approach to categorization which allows for different levels of action concepts, namely prototypical Scenes, Action Types, and Metacategories.

1.2 Scenes as prototypes for action concepts

The development of Scenes is the final step of the IMAGACT ontology-building process. Up to that point this process has been developed through the manual annotation and classification of action verbs retrieved from large spoken corpora of Italian and English (for a detailed account of this procedure, see Moneglia et al., 2012a; Moneglia et al., 2012b). In the Scene creation step, action classes are demarcated on the basis of semantic differentials between the verbs. Each action class is then linguistically motivated by the presence of a unique set of Italian-English verbs that can be used to refer to it.

To this end, IMAGACT made use of systematically annotated Local Equivalence phenomenon, i.e. the possibility that different verbs, with different meanings, refer to the same action class (we will elaborate on this in Section 2).

For instance, if *someone takes something off the floor*, we could also say that *someone picks something up*: this means that between *to take* and to *pick up* there is a Local Equivalence in this specific field of application. On the contrary, this relation is not valid for the action described by the sentence *someone takes something from a (high) shelf*, which is not a possible extension of the verb *to pick up*. Since we can apply both of these verbs to the first event, but only *to take* to the second one, we have discovered a linguistic differential between these two action classes: this fact led to the production of two different Scenes.

[1] Freely accessible at http://www.imagact.it/

This procedure ensures a good definition of action identification, which cannot be only function of the verb thematic structures (as in VerbNet; Kipper-Schuler, 2006). For instance, the sentences *he takes/get the water*, *he takes/grasp the handle* and *he takes the glass* show the same thematic structure, but refer to different actions, as the differential in Local Equivalence testifies.

Finally, a prototypical action has been chosen for each class, and represented by a recorded video or 3D animation. The IMAGACT database contains 1,010 Scenes, which constitute the basic entities of reference of the action ontology, linked primarily to the English and Italian verbs considered in the annotation (more than 500 for each language). After this bootstrapping process, the ontology was extended to many other languages[2] via competence judgments given by native speakers for each Scene[3] (Brown et al., 2014; Pan, 2016).

This way, the set of Scenes to which a verb is connected is, in fact, a sampling of the unlimited possible actions referred to by that verb. Moreover, the IMAGACT methodology ensures this sampling to be representative of the whole semantic variation of each verb.

Aside from all this, the problem of the identification and formalization of the action concepts still remains. A great number of linguistic differentials may occur within the range of the most general action verbs, which are also some of the most frequently occurring; for example, the verb *to take* refers to more than 100 IMAGACT Scenes.

Table 1 reports the number of verbs connected to the Scenes. In order to have a readable picture, 5 groups have been identified with respect to the verb generality degree: verbs connected to more than 30 Scenes (i.e. very general verbs, that can be used to refer a wide variety of different actions), to 11-30 Scenes, to 5-10 Scenes, to 2-4 Scenes, and to 1 Scene only (i.e. very specific verbs). Values are reported in percentage on the total number of the annotated verbs of each language[4].

	>30 s.	**11-30 s.**	**5-10 s.**	**2-4 s.**	**1 s.**
Arab	0.7%	**5.7%**	**16.1%**	**38.5%**	39.0%
Chinese	0.0%	0.2%	1.9%	19.3%	**78.5%**
Danish	0.2%	3.3%	8.8%	27.7%	**60.1%**
English	**1.3%**	**5.4%**	**17.7%**	**40.1%**	35.6%
German	0.0%	2.2%	6.3%	30.0%	**61.5%**
Hindi	0.4%	3.3%	7.2%	24.4%	**64.6%**
Italian	**1.1%**	**5.6%**	**18.0%**	**37.4%**	37.9%
Japanese	0.0%	2.1%	8.7%	28.6%	**60.6%**
Polish	0.0%	1.6%	10.0%	32.0%	56.4%
Portuguese	**1.1%**	**5.8%**	10.8%	30.1%	52.2%
Serbian	0.2%	2.5%	8.3%	30.6%	58.4%
Spanish	**1.1%**	**5.8%**	10.9%	33.0%	49.2%

Table 1: Percentage of verbs linked to the Scenes

A clearer picture of this phenomenon is shown in Table 2, reporting the percentage of verb-scene relations; it can be read as a measure of the impact that general vs. non-general verbs have in action categorization for each language. For example, according to Table 1, English verbs that can be considered very general are 1.3% of the annotated verbs, but they are involved in 16.8% of the whole set of verb-scene English relations (Table 2).

	>30 s.	**11-30 s.**	**5-10 s.**	**2-4 s.**	**1 s.**
Arab	10.6%	**23.6%**	**29.7%**	26.0%	10.0%
Chinese	0.0%	3.3%	9.1%	**33.6%**	**54.0%**
Danish	2.2%	22.1%	22.9%	28.8%	24.1%
English	**16.8%**	20.9%	**28.8%**	25.1%	8.5%
German	0.0%	17.2%	19.5%	**34.3%**	**29.0%**
Hindi	5.8%	21.1%	20.6%	25.4%	**27.1%**
Italian	**14.1%**	**23.5%**	**29.3%**	23.8%	9.3%
Japanese	0.0%	14.7%	25.6%	**33.0%**	26.7%
Polish	0.0%	9.9%	**28.7%**	**36.3%**	25.1%
Portuguese	**17.1%**	**26.6%**	20.0%	21.8%	14.4%
Serbian	4.4%	15.5%	22.8%	**33.3%**	24.1%
Spanish	**16.8%**	**26.6%**	19.5%	23.9%	13.3%

Table 2: Percentage of verb-scene relations

Tables 1 and 2 clearly show that different languages adopt different lexicalization strategies to refer to the action universe. For instance, general verbs are preeminent in romance languages and in English (the impact of verbs linked to more than 10 Scenes is above 35%), while Chinese has the lowest presence of general verbs and the highest impact of verbs connected to only one Scene (54%).

1.3 Higher levels of conceptualization: Types and Metacategories

In order to identify higher level action concepts within the broad range of prototypes representing a verb's variation (e.g. the ones in Figure 1), we need to make clusters of conceptually similar Scenes. This step is also needed to give a cognitively plausible account of their semantic variation with a reasonable level of granularity.

Similarity judgments among Scenes could help to gather action classes into broader sets, but how is this possible in practice? Moreover, verb semantics strongly influences these similarity judgments: even if two action classes show a linguistic differentials, they can appear conceptually similar if we look at them from the perspective of a very general verb. For instance, the two above-mentioned actions of *taking something off the floor* and *taking something from a shelf* can be considered within the same, wider, action concept if we look at them from the perspective of the verb *to take*, in which case the linguistic differential of *to pick up* is somewhat irrelevant.

Action Types in IMAGACT are defined as action concepts within the semantic variation of a verb. The creation of Types was performed independently of each other in the Italian and English corpora by mother tongue annotators through a corpus-driven process of associating similar actions. The set of Types for each verb is in fact a segmentation of its semantic variation where each Type is represented in the IMAGACT ontology as a clustering of Scenes.

At a higher level of conceptualization, the numerous actions covered by IMAGACT have been gathered into 9

[2] A further 10 languages are completely mapped (see Tables 1 and 2) and 16 are under development.

[3] The competence judgments were recorded through a dedicated web interface. The interface shows the native speaker a scene and they are asked to answer the question: *how can you say this action in your language?*

[4] The number of annotated verbs is very different among the languages, from a minimum of 414 (Chinese) to a maximum of 1193 (Polish): this depends on linguistic differences among languages and not on the partial status of the annotation work, that is completed for these 12 languages.

Metacategories, characterized as typical of human categorizations of action. These metacategories are ordered according to criteria that take into account the informative focus of the action, as shown in Table 3.
In short, within the IMAGACT framework each action can be categorized in three ways: a) belonging to an action class represented by a Scene and linked to different verbs (in various languages); b) belonging to different Action Types; c) belonging to one (or in some cases two) Metacategory. Scenes, Action Types and Metacategories thus constitute conceptualization options with differing levels of granularity.

AGENT perspective	**AGENT-THEME relation**	**THEME-DESTINATION relation**
Actions referring to facial expressions	Modification of the OBJECT	Change of location of the OBJECT
Actions referring to the body	Deterioration of the OBJECT	Setting relations among OBJECTS
Movement in space	Force on the OBJECT	Actions in inter-subjective space

Table 3: Action Metacategories

2. The Role of Local Equivalence

As we already said, the main problem for the linguistic annotation of action concepts, both in language and scene datasets, is the identification of the entities that should constitute the reference points in the ontology of actions.
In this section and the subsequent ones we will show (abstracting away from the concrete implementation of these concepts in the IMAGACT resource) how the Local Equivalence can be exploited as a powerful annotation tool for action identification.
Insofar as one verb may refer to many actions, each action may also be identified through various lexical alternatives. We called this property Local Equivalence, since it is valid only within certain *local* application of the verbs, and it is not a property belonging to their (*general*) meaning. Local Equivalence, then, associated with the productivity of action concepts, can be used to reduce the underdetermination and the granularity of action concepts.
Looking at the variation of *to take*, almost every action prototype features one or more Local Equivalence relations with other action verbs, e.g. *to extract*, *to receive*, *to remove*, *to bring*, *to lead*, *to grasp*. Figure 3 shows a snapshot of the referential variation of the verb *to take*, re-organized in consideration of the abovementioned equivalences. These equivalences constitute explicit differences between each action concept prototype and the others, or, in another sense, a restriction of its boundaries.
The action concept grouping the scenes in the top left corner of the figure (labeled as *remove*) is split from the one on the right side (labeled as *bring*) because the former holds an equivalence between *to take* and *to remove*, while the latter demonstrates the equivalence between *to take* and *to bring*.
The parsing of the action continuum into a discrete set of ontological entities can be further objectified by crossing the data of the linguistic categorization. When two different action verbs demonstrate the same event type, then that event type should be somehow considered as an identifiable action concept. Local Equivalence provides for the parsing of action concepts as they are referred to in different languages.
Once the variation and differentials are identified, the action concepts can be modelled and generalizations obtained. As Figure 3 shows, the set of actions extended by *to take* fall into a restricted set of models roughly identified by their higher level Local Equivalences (specifically *to remove*, *to bring*, *to receive*, and *to grasp*). Within these broad concepts, we can refine the granularity

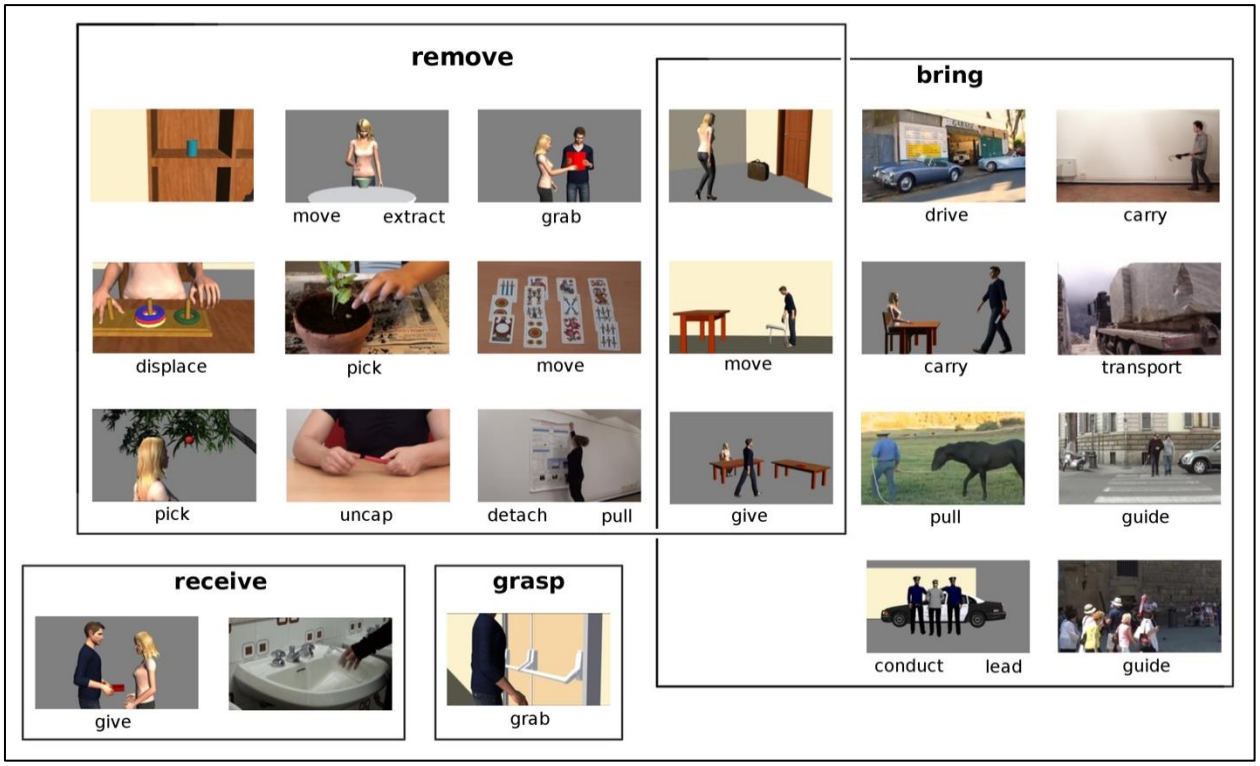

Figure 3: The referential variation of *to take* organized using Local Equivalence relations.

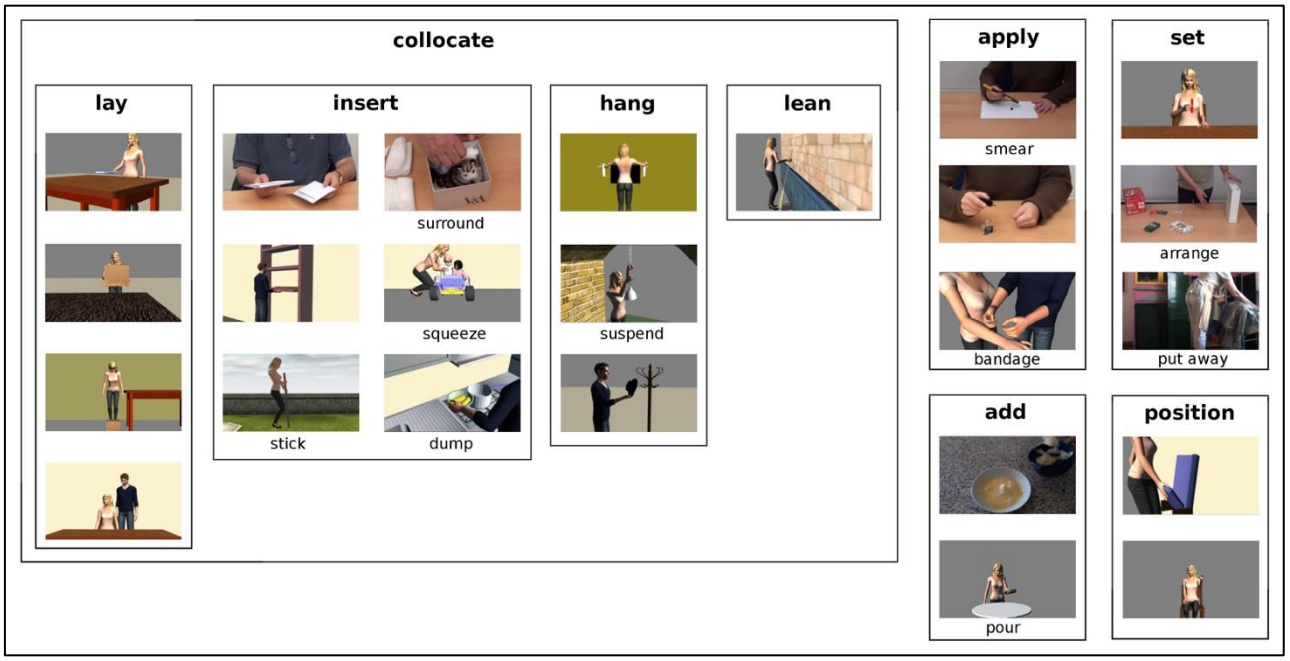

Figure 4: The referential variation of *to put* organized using Local Equivalence relations.

of the conceptualization using more specific equivalences (e.g. those with the verbs *to extract*, *to pick* or *to detach*).
This step opens up the path to identifying how languages vary in segmenting the action universe, as we have seen for *toru* in Japanese, whose range of variation shows an intersection with English that is observable through comparison.
With regard to hierarchical relations among action concepts, the Local Equivalences specify how lower-level and higher-level concepts are organized in the conceptual structure and how cross-categorization phenomena characterize the hierarchy. In Figure 3, for instance, *taking*/*removing* and *taking*/*bringing* events correspond to two hierarchies, which may intersect with *moving* and *giving* type events.
This framework has been applied extensively in IMAGACT for analyzing action verbs in many languages. Figure 4 shows the Local-Equivalence-based grouping of action class prototypes referred to by the general verb *to put*. Nevertheless, the framework we described asks for a stricter definition of Local Equivalence relations: how and why can two verbs extend to the same action concept? What are the limits of the application of Local Equivalence for the ends of action identification?
In the following Sections, we will try to disentangle the different phenomena underlying the observation of Local Equivalences, distinguishing among them with respect to their usability in the complex task of action concept identification.

3. Local Equivalence as a Function of Semantic Properties

Local Equivalence can be a function of verb semantics. Let's consider Figure 5, which is one of the prototypes in the variation of *to hang*. In that prototype, as in almost all prototypes in the variation, *to put* can also be applied. As a matter of fact, a competent speaker of English may refer to the event with both the sentences *John hangs the hat on the hook* and *John puts the hat on the hook*.
The reason for this equivalence relies on semantic factors, and it is not a result of occasional and pragmatic circumstances. Very roughly speaking, one could say that both actions (*to put* and *to hang*) have the same GOAL of giving a LOCATION to the hat (i.e. *to collocate*) and for this reason the two predicates record a Local Equivalence relation for these kinds of events.

Figure 5: *John hangs/puts the hat on the hook*
http://bit.ly/2HSk9Du

It should be clear that the Local Equivalence relation between *to hang* and *to put* with respect to this action class does not imply that the two abovementioned sentences (and verbs) have the same meaning. While the first one (containing the verb *to hang*) specifies the MANNER in which the hat is placed on the hook (i.e. it encodes a feature of the action's RESULTING STATE), the second sentence (with *to put*) does not: it simply specifies the LOCATION of the THEME. This is the reason why we do not treat Local Equivalence as a synonymy relation[5]. No synonymy occurs: quite simply, either verb may be substituted into the sentence maintaining the same reference, but not the same meaning (Frege, 1892).
The referential equivalence between the verbs *to put* and *to hang* is not restricted to the event represented in Figure 5, but instead extends to any action of the same class. Generally, whenever an AGENT places something in a LOCATION and its RESULTING STATE is "suspended", we can use both *to put* and *to hang* to refer to that action.

[5] Therefore, Local Equivalence relations are not suitable for creating synsets in a WordNet-like scenario.

More specifically, the possibility of applying the verb *to put* to this event type arises for two general reasons: i) if something hangs, then it must have a definite LOCATION from which it hangs; ii) an OBJECT can be considered a LOCATION at the conceptual level (see, for instance, Jackendoff, 1983). This means that, in this case, Local Equivalence is a productive relation.

Being productive for semantic reasons, Local Equivalence determines the identification of an action concept and distinguishes it from the other fields of application of both verbs where this specific relation does not occur. The action concept identified constitutes a conceptual entity through which we can categorize the actions falling within the extensional variation of *to put*.

Within this variation, it's possible to identify a set of troponymic concepts that are based on the quality of the RESULTING STATE of the THEME, as the ones represented in Figures 5, 6, and 7. In all of these cases we have a specific Local Equivalence (respectively *to put*/*to hang*; *to put*/*to lay*; *to put*/*to lean*) that is productive and relies on semantic factors. This fact presents a linguistic motivation for categorizing these events as three different action concepts to which we can refer with *to put*.

It's important to stress that these relations between verbs exist only locally and cannot be extended to a more general lexical level. The LOCATION of the THEME, for example, occurs in almost all variations of *to hang*, but the feature "reaching a LOCATION" is not strictly necessary for the eventualities in the extension of this verb. In particular, *to hang* also records interpretations in which no locative event occurs, like *Mary hangs her head* in Figure 8. Similarly, there are many instances of *putting* events where the RESULTING STATE is not "suspended", as we see for the examples in Figures 6 and 7.

Figure 6: *Mary puts/lays the book on the table*
http://bit.ly/2FcaKb4

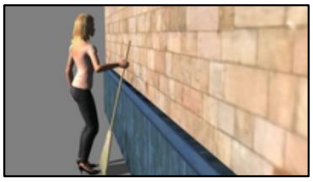

Figure 7: *Mary puts/leans the broom against the wall*
http://bit.ly/2FB0QOe

Figure 8: *Mary hangs her head*
http://bit.ly/2GReR9P

Once again, we have to underline that Local Equivalence is properly *local* because it does not allow the induction of entailments or other semantic relations at a lexical level: general statements like *if I put then I hang* or *if I hang than I put* are false. Instead, a relation between verbs is valid within the scope of a specific, identifiable action concept.

4. Local Equivalence as a Function of Productive Pragmatic Properties

Local Equivalence relations may depend on pragmatic factors only, but despite this fact their identification can still have huge consequences for the definition of action concepts. Let's consider the relation between the concepts of *taking* and *removing*. There is nothing in the meaning of *to take* which refers to the concept of DISPLACEMENT. The GOAL of *to take* has something to do with "getting something in the AGENT's control", and does not refer to "moving something from its previous LOCATION". In other words, it is not possible to predicate the action of removing something from its position with *to take*. However, looking at the events in which *to take* applies we see that, for many actions falling within its variation, when the AGENT takes the OBJECT under his control, the OBJECT also loses its original LOCATION (see Figure 9). Interestingly, this does not happen in cases where *to take* is equivalent to *to grasp* (Figure 10) or *to receive*.

Figure 9: *John takes/removes the cup from the shelf*
http://bit.ly/1eoMuOW

By consequence, *to take* records a Local Equivalence relation with *to remove*. This equivalence does not occur by chance, and is a direct consequence of the following pragmatic circumstance: if we get something in our possession, this causes the DISPLACEMENT of the object. In other words, this correlation is pragmatic, but not occasional, and corresponds to the systematic equivalence of the two verbs in most of the semantic variations of *to take*.

The consequences of the annotation of this Local Equivalence in defining the identity of the set of action concepts which fall in the variation of *to take* are important. The property of DISPLACEMENT and the parallel Local Equivalence relation with *to remove* is a relevant feature of certain action concepts falling under its variation and is not represented in the meaning of the verb.

This relevance is provable through similarity judgments: if the equivalence is lost, then the action is perceived as belonging to a different class. For example, if the AGENT reaches for a cup and grasps it without moving it, the action falls into the action type of *grasping*, represented in Figure 10. In the opposite case, if the AGENT in Figure 10 grasps the bar and removes it from the door, the action is judged as similar to *taking the cup*.

The pragmatic aspect of OBJECT DISPLACEMENT is a differential feature for a set of action concepts in the variation of *to take*, though it is not a semantic feature of the language concept.

Figure 10: *John takes/grasps the handle*
http://bit.ly/1ftSeCC

5. Local Equivalence and Co-Occurrence for Different Actions

Interpersonal activities are relevant to human categorization and they constitute one of the basic stages in the cognitive development of the child (Tomasello 2009). Events that are the product of these activities are by necessity composed of various synchronous actions performed by the participants. Therefore, the verbs referring to those activities end up being equivalent for the identification of that event. IMAGACT records these actions under one specific action Metacategory (see Table 3). For instance, the verb *to take*, when referring to a frame dealing with intersubjective activity, specifies an action type in which *taking* something is synchronous with the activity of *receiving* the object, and with an act of *giving* performed by the second actor in the intersubjective action. In this kind of event, the two actors co-operate and their activities are both necessary and synchronous with the onset of the concept.

The Local Equivalence relation between the two properties (*taking*/*receiving* and *giving*) is pragmatic, and is not represented in the meaning of *to take*, which does not require intersubjectivity. However, reference to this property is necessary to identify the variation of the referred action concepts. Specifically, if we want to distinguish *Mary takes the cup from the shell* from *Mary takes the cup from John (who gives it to Mary)*, the identification of the Local Equivalence between *to take* and *to receive* constitutes a necessary annotation.

6. Nonproductive Pragmatic Equivalences

The onset of Local Equivalence relations that follow from pragmatic factors is pretty frequent when working with prototypes with the aim of representing action concepts, however in many cases Local Equivalences are not relevant for the identification of these concepts.

For instance, among the action types in which *to take* is equivalent to the verb *to lead* there is the event represented in Figure 11, in which a Local Equivalent relation with the verb *to guide* is productive. Beyond this equivalence, which distinguishes this action concept from the others in the variation of *to take*, the prototype also represents the synchronous action of *crossing* the street. This property is prominent in the prototype, and the two concepts (*taking*/*leading*/*guiding* and *crossing*) are also frequently associated in the world when people need to be guided, since *crossing* is a difficult task for them.

Therefore, the Local Equivalence among these 4 verbs is noticeable in that prototype, and the event can be properly described with both the sentences *John takes/leads/guides the blind man across the street* and *John and the blind man cross the street*.

Figure 11. *John takes/leads/guides the blind man across the street*; *John and the blind man cross the street*
http://bit.ly/X9aoZj

The event is therefore an extension of both *to guide* and *to cross*, but it is worth noting that the Local Equivalence provided by *to cross* does not contribute to the identification of this action concept. Indeed, a modification to the prototype which discards the property of *crossing* (e.g. *the blind man is guided along a street*) does not change the action type. In other words, the Local Equivalence between *to take* and *to cross* is not productive, while the one between *to take*, *to guide* and *to lead* is productive and identifies a concept within the variation of *to take*. More concisely, the property of *crossing* does not underly the concept of *guiding* and does not constitute a proper troponymic concept.

7. Concluding remarks

The problem of identifying action concepts can be (at least partially) solved through the annotation of the systematic co-referential properties of action verbs. Indeed, Local Equivalence phenomena delimit specific sectors in the action continuum, meaning that action concepts may be properly determined starting from linguistic categorizations.

Nevertheless, the annotation of Local Equivalences with the aim of identifying action concepts requires an evaluation of the productivity of the relation. Two actions are of the same type only if the concept extends in the same way, i.e. if they record the same productivity. When this productivity is missing the Local Equivalence is not essential and exists just as an accidental pragmatic fact.

This aspect yields an essential contribution to the annotation of action from a linguistic perspective: without considering the presence of Local Equivalence relations action concepts remain vague and strongly underdetermined and their categorization does not find adequate points of anchorage.

8. Bibliographical References

Brown, S.W., Gagliardi, G. and Moneglia, M. (2014). IMAGACT4ALL: Mapping Spanish Varieties onto a Corpus-Based Ontology of Action. *CHIMERA*, 1:91--135.

Fellbaum, Ch. (editor) (1998). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Frege G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25--50.

Jackendoff, R. (1983). Semantics and Cognition. Cambridge, MA: MIT Press.
Kipper-Schuler, K. (2005). Verbnet: A Broad-Coverage, Comprehensive Verb Lexicon. PhD Diss. University of Pennsylvania, Philadelphia, PA, USA.
Kopecka, A. and Narasimhan, B. (2012). Events of Putting and Taking, A Cross-linguistic Perspective. Amsterdam/Philadelphia: John Benjamins.
Majid, A., Boster, J. S., and Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, 109(2).
Miller, G.A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11): 39--41.
Moneglia, M. (2014). Natural Language Ontology of Action: A Gap with Huge Consequences for Natural Language Understanding and Machine Translation. In Z. Vetulani and J. Mariani, editors, *Human Language Technology. Challenges for Computer Science and Linguistics*. Berlin/Heidelberg: Springer, pp. 370--395.
Moneglia, M. (2014b). The variation of action verbs in multilingual Spontaneous speech Corpora. In T. Raso and H. Mello, editors, *Spoken Corpora and Linguistics Studies*. Amsterdam: Benjamin, pp. 152--190.
Moneglia, M., Gagliardi, G., Panunzi, A., Frontini, F., Russo, I. and Monachini, M. (2012a). IMAGACT: deriving an action ontology from spoken corpora. In *Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-8)*, pp. 42--47.
Moneglia, M., Monachini, M., Calabrese, O., Panunzi, A., Frontini, F., Gagliardi, G., and Russo, I. (2012b). The IMAGACT Cross-linguistic Ontology of Action. A new infrastructure for natural language disambiguation. In N. Calzolari et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association (ELRA), pp. 948--955.
Pan, Y. (2016). Verbi di azione in italiano e in cinese mandarino. Implementazione e validazione del cinese nell'ontologia interlinguistica dell'azione IMAGACT. PhD diss., University of Florence.
Panunzi, A., De Felice, I., Gregori, L., Jacoviello, S., Monachini, M., Moneglia, M., and Quochi, V. (2014). Translating action verbs using a dictionary of images: the IMAGACT ontology. In A. Abel, C. Vettori, and N. Ralli, editors, *Proceedings of the XVI EURALEX International Congress: The user in focus*. Bolzano: EURAC research, pp. 1163--1170.
Rosch, E. (1978). Principles of Categorization. In E. Rosch and B.B. Lloyd, editors, Cognition and Categorization. Hillsdale, NW: Erlbaum, pp. 27--48.
Rosch, E. (1983). Prototype classification and logical classification: The two systems. New trends in conceptual representation: Challenges to Piaget's theory, pp.73--86.
Tomasello, M. (2009). Why we cooperate. Cambridge, MA: The MIT Press.
Umiltà M.A., Escola L., Intskirveli I., Grammont F., Rochat M.J., Caruana F., Jezzini A., Gallese V., Rizzolatti G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of The National Academy of Sciences*, 105(6):2209--2213.

9. Language Resource References

IMAGACT. http://www.imagact.it/