

The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action

Moneglia, M.¹, Brown, S.¹, Frontini, F.², Gagliardi, G.¹, Khan, F.², Monachini, M.², Panunzi, A.¹

¹University of Florence

²ILC CNR

Piazza Savonarola 1, Florence (Italy)

Via Giuseppe Moruzzi, 1 Pisa (Italy)

E-mail: moneglia@unifi.it, susanwbrown@att.net, francescafrontini@gmail.com, gloria.gagliardi@gmail.com, anasfkhan81@gmail.com, monica.monachini@ilc.cnr.it, alessandro.panunzi@unifi.it

Abstract

Action verbs have many meanings, covering actions in different ontological types. Moreover, each language categorizes action in its own way. One verb can refer to many different actions and one action can be identified by more than one verb. The range of variations within and across languages is largely unknown, causing trouble for natural language processing tasks. IMAGACT is a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech corpora, which makes use of the universal language of images to identify the different action types extended by verbs referring to action in English, Italian, Chinese and Spanish. This paper presents the infrastructure and the various linguistic information the user can derive from it. IMAGACT makes explicit the variation of meaning of action verbs within one language and allows comparisons of verb variations within and across languages. Because the action concepts are represented with videos, extension into new languages beyond those presently implemented in IMAGACT is done using competence-based judgments by mother-tongue informants without intense lexicographic work involving underdetermined semantic description.

Keywords: Action ontology, Semantic variation, Visual representation

1. Introduction

In order to communicate, it's necessary to refer to actions. Speakers do not have a problem finding the right verb for a specific action in their own language. However, in a foreign language, they often have difficulty choosing the appropriate verb. The reason is that one verb can refer, in its own meaning, to many different actions and we cannot be certain that the same set of alternatives is allowed in another language (Kopecka & Narasimhan 2012).

Like second-language learners, automatic translation systems suffer from this problem, even when the translation of a simple sentence is required. For instance, asking Google to translate sentences containing action verbs from English to Italian and vice-versa results in systematic problems, as in the following examples:

“John pushes the garbage into a can”

the system returns

“John spinge l'immondizia nel secchio”

instead of

“John comprime l'immondizia nel secchio”.

Working in the other direction:

“Mario prende il gatto per la coda”

the system returns

“Mario takes the cat by the tail”

instead of

“Mario catches the cat by the tail”,

“Mario gira gli zucchini nella padella”

the system returns

“John turns the zucchini in the pan”

instead of

“John stirs the zucchini in the pan” .

The reason for these mistranslations is that the set of possible interpretations of general action verbs, such as *prendere*, *spingere*, *girare*, *comprimere* in Italian and *take*, *catch*, *push*, and *turn* in English and their cross-linguistic correspondences, are not mastered by the system and, indeed, are not explicitly settled in any current language resource (Moneglia 2011; Moneglia et al. 2012a).

The problem is a significant one because reference to action is very frequent in ordinary spoken communication, and these high-frequency verbs can each refer to many different action types (Moneglia in press).

The IMAGACT project has now delivered a multimodal language infrastructure covering the set of actions most frequently referred to in everyday language. Using English and Italian spoken corpora, we have identified 1010 distinct action concepts referred in the spoken language interaction and visually represented them with prototypical scenes, either animated or filmed. The cross-linguistic correspondences of those actions with the verbs that can refer to them in English and Italian have been established in a MySQL database through an annotation infrastructure, allowing the induction of action concepts from corpus data. The infrastructure and the annotation procedure have been presented in previous work (Frontini et al. 2012; Moneglia et al. 2012b).

This ontology gives a picture of the variety of activities that are prominent in our everyday life and specifies the language used to express each one in ordinary communication. 521 Italian verbs and 550 English verbs (i.e., the verbal lexicon most likely to be used when referring to action) are stored therein.

Each verb can express one or more concepts, while each concept can refer to one or more verbs (within and across languages).

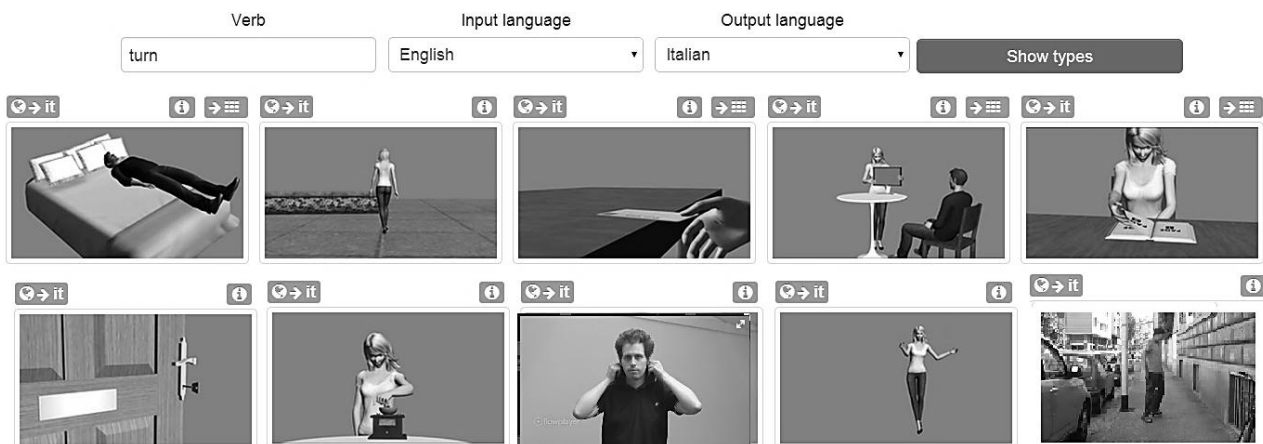


Figure 1: The variation of *to turn* across action types

IMAGACT is now freely available to the user on line. This paper presents the multimodal infrastructure and describes the responses that would be made to typical queries of the database.

Because the action concepts are represented in an ontology of animated and filmed scenes, they can be easily identified by speakers of any language, making the resource widely accessible. In the final sections, we present the resource’s dedicated collaborative framework and interface, which have been designed to validate the resource and facilitate IMAGACT’s expansion into other languages. The action concepts in IMAGACT have already been extended through this method to Chinese and Spanish (included in the first IMAGACT release).

2. From Language to Scenes and from Scenes to Language

IMAGACT can be used as a multilingual dictionary of images. For each of the four languages included in the IMAGACT first release, the interface makes explicit through short videos the range of variation of each action verb across the different action types it can refer to. On this basis the infrastructure also specifies the cross-linguistic correspondences that are possible for each represented action. Figure 1 show the thumbnail images of the main types of actions identified by the English verb *to turn*.

Looking at the various action types this verb expresses in English, the user can select the action type he is interested in, watch the animation clip to clarify the meaning, and see how this action is identified in the target language, for instance Italian. IMAGACT returns one main verb and an additional set of verbs that equally identify this specific type of action.

Figure 2 shows how the system returns the information for two action types. In almost all cases for the verb *to turn*, the user will see that the appropriate entry in Italian is the verb *girare*, as in the case of the postcard.

However, among the set of actions referred to by *to turn*, the user may be interested in the one where *turn* is equivalent to *fold*. IMAGACT reveals that this action requires the use of a different verb in Italian (*rigirare*).

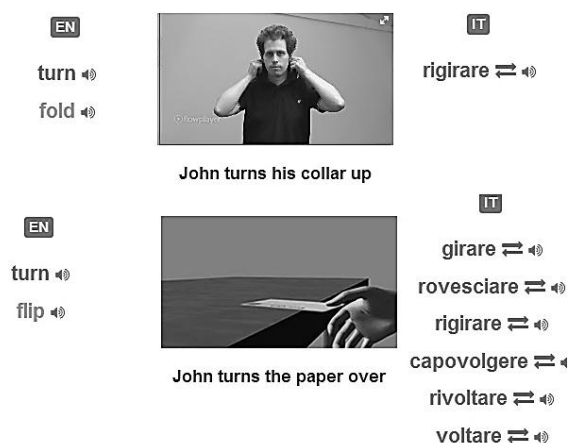


Figure 2: The cross-linguistic relation of verb(s) to action type

3. Comparison

IMAGACT allows the comparison of verbs from two different languages that in principle should translate between each other. Searching with this function, the system illustrates the set of domains in which both verbs can be applied. The result of such a search for *to turn* and *girare* (Figure 3) supports the intuition that the two verbs can translate to each other. At the same time, however, the system shows which actions can be indicated by one verb but not by the other, and vice versa. As a consequence, the difference between the Italian verb *girare* and the English verb *to turn* becomes explicit.

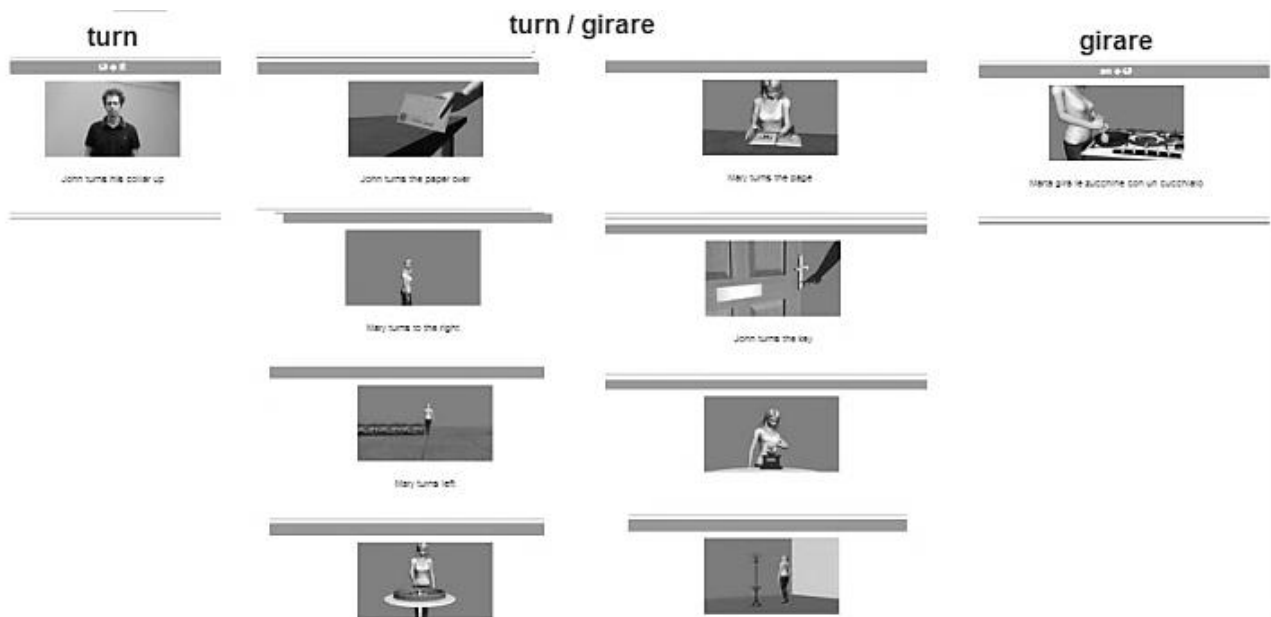


Figure 3: Comparison of *turn* vs. *girare*

English and Italian users can verify how the range of variations allowed by their language concepts over-extend or under-extend the variation allowed in the target language. The Italian user will learn, for instance, that in English *turn* cannot be applied to the type in Figure 4. The verb *stir* should be used instead.



Each action type among those represented by a particular verb is also related to a set of equivalent verbs. Comparisons within language allow the user to explore more deeply the differences in meaning between the lexical entries suggested by the system. For instance, an English user focusing on the case in Figure 5 may wonder what the difference is between the two Italian lemmas suggested by the system (*girare* / *mescolare*).

To answer this question, the user can compare two verbs of the target language (Figure 5). By considering the variation of the two verbs, the user can figure out how their meanings differ and the range of each verb's usage. *Mescolare* refers more generally to events in which things are mixed.

Figure 4. From comparison to alternative categorizations

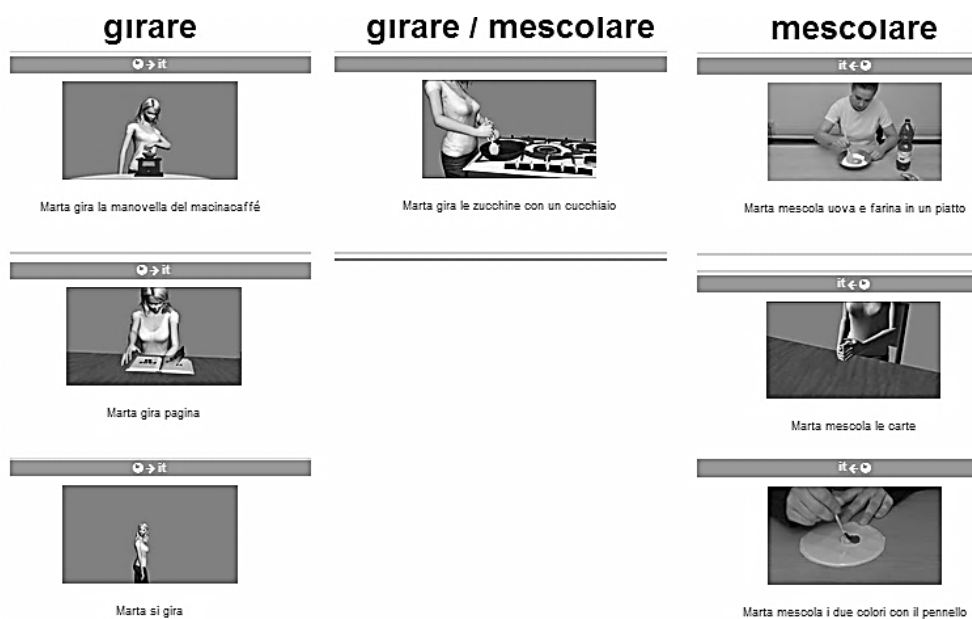


Figure 5: Intra-linguistic comparison

4. Gallery

If the language of the user is not represented in IMAGACT, she can use the system directly as a gallery of scenes. This may be of special interest to users who speak minority languages.

The system works similarly to the scale or the self-checkout machine at the supermarket that allows the shopper to select their item from images of fruits and vegetables. Categories of actions have been represented through meta-animation in 3D, which are played in loop, as the thumbnails in Figure 6 suggest.

The user can figure out what these classes represent by looking at the abstract representation heading each class and of course through a quick look at the actions gathered under each one. The process that leads a user to identify the action she is interested in and its linguistic categorization in a target language is independent of the word she has for that action in her language. From this point of view, IMAGACT reverses the ordering of the dictionary; i.e. from concepts to language instead of from language to concepts.

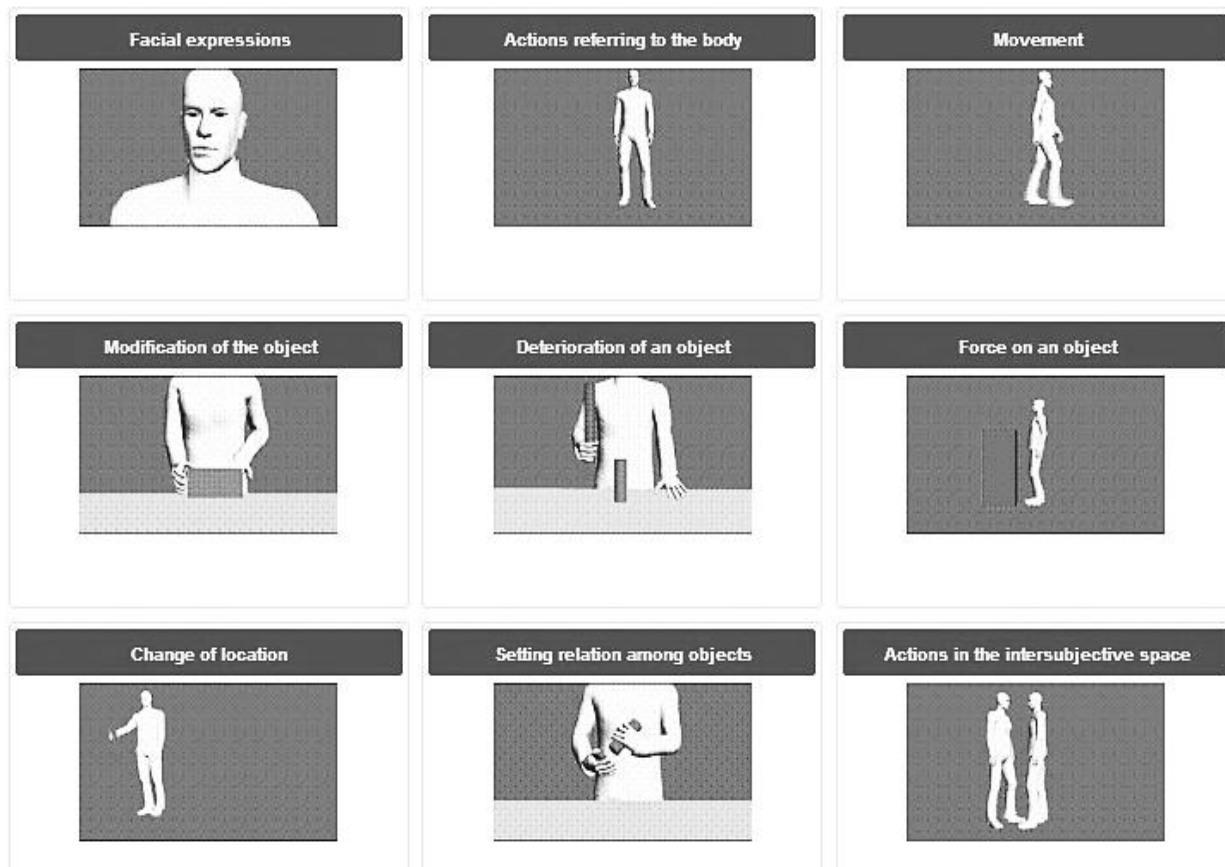


Figure 6: The gallery of meta-categories

The infrastructure gathers the numerous actions covered by IMAGACT into 9 classes, which have high relevance in human categorization of action.

Categories are ordered according to criteria which follow the informative focus of the action:

Perspective centered on the Actor:

- Actions referring to facial expression
- Actions referring to the body
- Movement in space

Perspective centered on the Actor-Theme relation:

- Modifications of the object
- Deterioration of the object
- Forces on the object

Perspective centered on the Theme-Destination relation:

- Change of location of the object
- Setting relations among objects
- Actions in inter-subjective space

Once the user has understood the meaning of the action groups, it will be easier to search for the specific action she is interested in. She will click on one scene in the gallery headed by one category and get the linguistic categorization of the concept in one of the possible target languages in IMAGACT.



Figure 7: From gallery to linguistic categorization

For instance, Figure 7 is what the system returns when asked for the Chinese verb for the action corresponding to *turn* under the category *Intersubjective space*.

5. Categories of Action as a Cognitive Correlation of Action Types

The corpus-based strategy pursued in IMAGACT for bootstrapping action concepts from spontaneous speech data relied on an induction process that separated the metaphorical and phraseological usages from occurrences referring to physical action and then classified the action occurrences into types, keeping granularity to its minimal level. (Moneglia et al. 2012a; Moneglia et al. 2012b). All action concepts in the ontology are physical actions extended by verbs in *their primary meaning*. This ensures, in principle, the productivity of each concept.

However, the induction and the representation of action concepts from the actual reference to physical actions found in corpora is a complex task, whose results can be corroborated at various levels.

The positive selection of occurrences in which verbs refer in their own meaning to physical actions preceded the annotation and constitute the milestone to bootstrap genuine Action concepts from large language corpora, in which, according to our finding, around 50% of occurrences are metaphors or idioms (see tables distributed in the IMAGACT project web site). The agreement on this task is high (9.5 K-Cohen agreement between experts reported in Gagliardi, 2014).

The annotation procedure, which has been presented in other works (Moneglia et al. 2012b), leads from occurrences of each verb in a language corpus to the identification of the possible action types productively referred to by the verb.

The possible variation of each language verb found in the corpus has been made explicit by gathering occurrences under prototypes that have been represented in scenes filmed or animated .

Although the agreement on this task have been positively evaluated (8.2 K agreement reported in Gagliardi 2014), the relevance of the induced categories can be also proved from both a cognitive and a linguistic point of view.

In IMAGACT, Categories of actions are metadata of Action types. The assignment of each Action type to Categories has been accomplished by a supervisor through Scene metadata interface available in the annotation infrastructure. The work consists in the evaluation of the information focus presented by each scene according to the tagset and have been performed considering scene by scene. Each scene is assigned to one Category with no reference to the variation of one specific verb. In the event of possible multiple categorization one scene is assigned to two Categories.

As a consequence of this work the variation of a verb across different action types can be presented to the user also referring types to the Categories of action they have been assigned (Figure 8).

Once types extended by a verb are gathered under different Categories, the reason why the variation of a verb is conceived as a discrete set of types instead of to a continuous variation become more clear. For instance the action types of a general verb like *turn* range over many different categories of action and Figure 8 shows how they are distributed.

The activities that are called “*turn*” range over six different action Categories, from *movement in space* to *intersubjective activities*. It is also for this reason that annotators found reasonable to distinguish the variation of this verb into Action types.

The division of the Action ontology into Categories of actions provides an independent cognitive correlations to the classification of action verbs variation into different types and, from the point of view of the user make clear how the variation of a single action verb in its own meaning systematically range over activities belonging to different domains of the human/world relation.

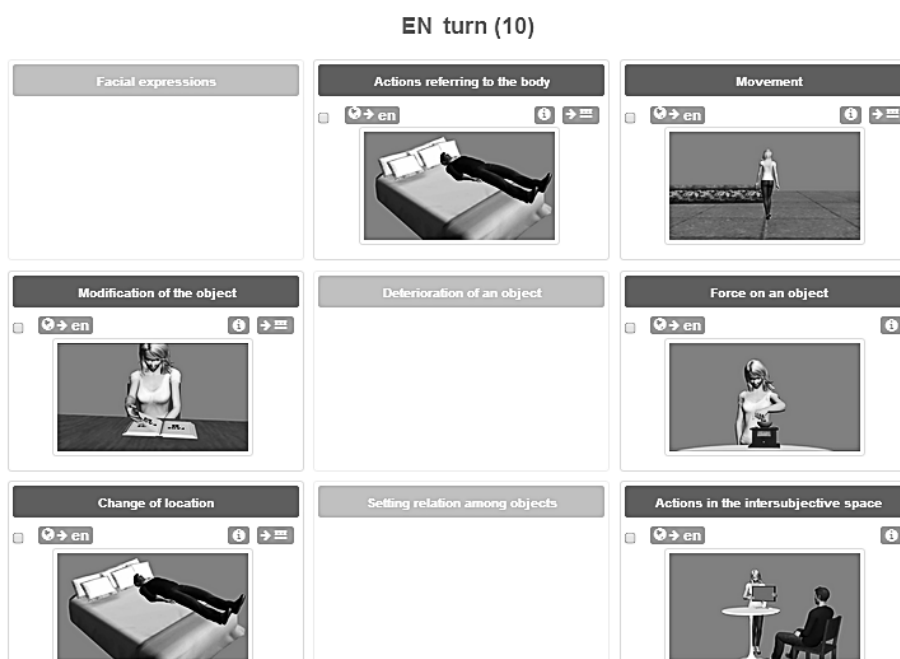


Figure 8: Viewing action types through categories (advanced user interface)

6. Competence-Based Validation and the Extension of the IMAGACT Ontology

6.1 Validation and productivity of concepts

Competence-based validation corroborates the linguistic productivity of types identified by the prototypical scenes in the IMAGACT ontology. In principle, a concept is valid for cross-linguistic reference to action if it is productive at the cross-linguistic level; that is, when, regardless of the language, the verb applied to the prototypic instance can also in principle be applied to the possible instantiations of the concept (Moneglia, 2011).

In IMAGACT the “instances of the concept” are represented by the set of sentences derived from corpus analysis gathered under one prototype. The interface allows one to verify whether or not the corpus-induced concepts fit with the language of the informant, indicating the extent to which the concept is productive at a cross-linguistic level.

We judge that an action concept in the ontology is fully productive at a cross-linguistic level (i.e., is a concept independent of the language) if all instances of that concept found in one language corpus that are categorized with one specific verb can also be categorized by at least a single verb in a different target language.

	★ to turn	★ to flip
	✕ ⇒	⇐ ✕
Sara gira la carta [della donna di cuori]	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Sara gira il cartoncino (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Matteo gira la cornetta (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Fabio gira la fotografia (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Il nutrizionista gira la scheda (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Cristina gira la cartolina (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Maria gira la cartolina (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Fabio gira la audiocassetta (4)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Il medico gira il bambino (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Sara gira la carta (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Sara gira la carta [della donna di cuori] (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Fabio gira la cassetta (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Il fotografo gira la foto (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Fabio gira la carta (2)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Cristina gira il dado (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Matteo gira il pettine (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>
<input type="checkbox"/> Matteo gira il libro (3)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input type="checkbox"/>

Figure 9: Proof of productivity IT > EN

This validation has been performed for the IMAGACT project by ILC for all the action types in the extension of each verb of the English and Italian corpora and has resulted in a measurement of the extent to which the 1010 concepts are equally productive for both English and Italian (Monachini, 2013).

The objects of the validation process are the source

verb, its action types, the scene representing the prototypical instance of each type and the standardized sentences related to each prototype in the source language.

An expert annotator who is a native speaker of the target language chooses the most suitable verb (or verbs) in his language to refer to the action represented by the 3D scene. The infrastructure then presents the set of source-language sentences from the corpus that have been assigned to that scene and asks whether or not the new target-language verb can be applied to all sentences in the set. For example in Figure 9, which is linked to the third type of Figure 1, the annotator judges the application of the target verbs *to turn* and *to flip* to the Italian sentences with the verb *girare* gathered in that type. The type is fully productive for English since all instances can accept the same verbs.

Table 1 summarizes the results of this validation procedure for the English and Italian annotation. The native-English annotator judged that the English verbs mapped to an action concept were appropriate translations for 98.50 percent of the Italian sentences annotated with that type. Similarly, the Italian annotator judged that the Italian verb assignments matched the English sentence sets for the types 98.87 percent of the time.

	Verbs	Instances	Basic Types	Validated Instances
IT > EN	521	21483	1100	98,50 %
EN > IT	550	15474	1180	98,87 %

Table 1: Summary of Italian and English validation

6.2 Validation and extension to language-specific concepts

The IMAGACT ontology was constructed with analyses of English and Italian corpora; therefore, it is possible that a new language will reveal a slightly different demarcation of concepts.

Because IMAGACT’s direct representation of actions through scenes can be interpreted independently of language, the infrastructure allows the mapping of lexicons from different languages onto the same cross-linguistic ontology. On this basis, it is possible to ask mother-tongue informants which verb(s) in their languages should be applied to each scene, thus extending the ontology to any language. IMAGACT has delivered an infrastructure that allows

- validation of the productivity of the actual concepts and implementation of the represented concepts according to a new language;
- mapping the verbal lexicon of any language onto the existing ontology through competence-based judgments (CBE-Light, see next paragraph).

The implementation of the set of action concepts in IMAGACT can be achieved using the corpus data stored in the database. These data from which the ontology originates are exploited in a *CBE interface*, which can be

used to validate the cross-linguistic productivity of the IMAGACT prototypes beyond English and Italian through the same system used for the IT/EN EN/IT validation. The lexical competence of a mother-tongue informant of a target language will be deeply challenged in this effort. Table 2 shows the results from matching Chinese onto the Italian set of instances.

	Verbs	Instances	Basic Types	Validated Instances
IT > CN	505	16579	923	82,70 %

Table 2: Summary of validation for Chinese

Figure 10 is an examples of this verification accomplished with Chinese, which replicates for the Chinese language the proof of productivity just seen for English with the third prototype of *turn/girare* in Figure 1. Figure 11 shows the proof of productivity of a different type (second prototype in the second row of Figure 1).

The Chinese informant verified that the two scenes require two different verbs (*fān* and *zhuǎn*). However, in the first case, the verb *fān* was appropriate for all occurrences gathered in the corpus under this prototype. The first concept is therefore fully productive in Chinese.

Sara gira la carta [della donna di cuori] (0)	
	★ fān ✕ →
Sara gira la carta [della donna di cuori]	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Sara gira il cartoncino (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Matteo gira la cornetta (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Fabio gira la fotografia (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Il nutrizionista gira la scheda (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Cristina gira la cartolina (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Maria gira la cartolina (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Fabio gira la audiocassetta (4)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Il medico gira il bambino (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Sara gira la carta (1)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>
Matteo gira il libro (3)	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/>

Figure 10: Proof of productivity IT > CN

This is not the case with the second prototype. As Figure 11 shows, instances gathered under this type in the Italian annotation comprise both cases in which the Actor is animate (typically a person) and cases in which the Actor is a Cause (like the *wind* or a *machine*). This difference is recorded at the level of thematic structure but was not considered relevant to induce two prototypes. In other words, it has been assumed that *John turns the gear* and

the steam turns the gear are instances of the same concept. As the figure shows, Chinese requires a different categorization in this second case. Therefore, a new verb is added and the first one is rejected for a set of instances. The language-specific categorization can then be introduced in IMAGACT. Through this method, it is possible to exploit the appropriate categorization of the concepts settled in the ontology in any language.

This validation process can lead to the implementation of new action concepts in the ontology according to language-specific constraints.

	★ 转 zhuǎn ✕ →	★ 使..转动 shǐ...zhuǎn dòng (Phr.) ← ✕
Matteo gira la manovella [dell'ingranaggio]	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Matteo gira la manopola [dell'ascensore giocattolo] (4)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Matteo gira la manopola (3)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Matteo gira la manovella [della pompa di benzina giocattolo] (1)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Matteo gira la manovella (8)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Matteo gira la manovella [dell'ingranaggio] (1)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Cristina gira la manovella (1)	Y: <input type="checkbox"/> N: <input type="checkbox"/>	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/> P: <input type="checkbox"/>
Il vapore gira l'ingranaggio	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input checked="" type="checkbox"/>
L'ingranaggio gira le macchine (1)	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input checked="" type="checkbox"/>
Il vapore gira un perno (1)	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input checked="" type="checkbox"/>
Il vapore gira l'ingranaggio (1)	Y: <input type="checkbox"/> N: <input checked="" type="checkbox"/>	Y: <input checked="" type="checkbox"/> N: <input type="checkbox"/> P: <input checked="" type="checkbox"/>

Figure 11: Proof of productivity IT > CN

6.3 IMAGACT4ALL: The *Light* Competence Based Extension Interface

In a simplified interface (*CBE light*) the work for implementing languages into IMAGACT is direct. In this framework the set of action concepts represented by the IMAGACT prototypic scenes is assumed as a fixed-reference universe.

An informant receives the action types as input. Figure 12 shows the interface the informant would use for processing one action type and how this has been done in the case of Chinese. The interface presents the informant with the scene prototype and the matching English and Italian verbs derived from corpus analysis. The informant assesses the action represented in the video and provides the verb or verbs in his language that can be used to refer to that specific action.

Lemmas are annotated in infinitival form, as it is commonly reported in dictionaries, in the box corresponding to his language. For each lemma he then writes in the caption box a simple sentence in the present tense, filling all the arguments of the verb that properly describes the action. This sentence will be used as the caption of the scene in the language of the informant.

Both the verb and the caption should be written in the current writing system of the language of the informant. If this system does not use roman characters,

Corpus verbs	Type	Lang.	Caption		
girare	PRO	🇮🇹	Marta gira la roulette		
revolve	PRO	🇺🇸	Mary revolves the wheel		
spin	PRO	🇺🇸	Mary spins the wheel		



Verb	Transliteration	Rejected	Lang.	Caption	Transliterated caption
转	zhuàn	<input type="checkbox"/>	🇨🇳	李娜转轮子	lǐ nà zhuàn lún zi
转动	zhuàn dòng	<input type="checkbox"/>	🇨🇳	李娜转动轮子	lǐ nà zhuàn dòng lún zi

Figure 12: IMAGACT4ALL annotation interface

the informant also provides the verb and its caption in roman characters, as can be seen for Chinese.

Given that verbs with different meanings can identify the same action, the informant is asked to find multiple lemmas allowed by her language for each action. However, simply viewing one film may be not sufficient to elicit all alternatives. The infrastructure provides one simple means to stimulate the thinking of the informant. More specifically, corpus-based annotation generated English and Italian alternatives that fit with the represented scene. These verbs will function as suggestions for figuring out alternatives in the language of the informant. Therefore, after the first lemma has been determined, the annotator is requested to judge whether or not the alternatives suggested have translations in her language that can be used in referring to the event in question. If so, she will report a new verbal lemma and a new caption by adding a line to her language options.

The work of the informant must be supervised by a mother-tongue expert linguist before the language is mapped onto the IMAGACT data base. More specifically, an annotation can be rejected by the supervisor during revision if considered inappropriate.

Spanish and Chinese have already been implemented through IMAGACT4ALL, and various initiatives are currently being pursued to implement both minority languages and languages with high impact at the world level.

7. Conclusion

The key innovation of IMAGACT is distinguishing the identification of action types from their definition. Short videos specify the variation of natural language verbs and constitute the entries of a language-independent ontology of action. IMAGACT can be exploited for both language learning and to design systems that allow disambiguation of action verbs in different languages. Thanks to the universal language of images, extension into other languages is done without intense lexicographic work involving underdetermined semantic description. In addition, network effects promise that the more languages that are implemented, the greater the rewards, thus validating in the real world the significance of the concepts listed in the ontology.

8. Acknowledgements

The IMAGACT project is funded by the PAR/FAS program of the Tuscan region in Italy.

9. References

- IMAGACT. <http://www.imagact.it>
- Frontini, F., De Felice, I., Khan, F., Russo, I., Monachini, M., Gagliardi, G. & Panunzi, A. (2012) Verb Interpretation for Basic Action Types: Annotation, Ontology Induction and Creation of Prototypical Scenes. Paper presented at the CogAlex-III Workshop as part of the COLING 2012 conference. Mumbai (India), December 2012.
- Gagliardi, G. (2014). Validazione dell'Ontologia dell'Azione IMAGACT per lo studio e la diagnostica del "Mild Cognitive Impairment" (MCI). PhD Dissertation, University of Florence
- Kopecka, A. & Narasimhan, B. (2012) *Events of Putting and Taking, A Cross-linguistic Perspective*. Amsterdam: Benjamins.
- Monachini, M. (2013). IMAGACT validation report. Project report.
- Moneglia, M. (2011). Natural Language Ontology of Action. A gap with huge consequences for Natural Language Understanding and Machine Translation. In Z. Vetulani (ed), *Human Language Technologies as a Challenge for Computer Science and Linguistics*. November 25-27, 2011, Poznań, Poland.
- Moneglia, M., Monachini, M., Calabrese, O., Panunzi, A., Frontini, F., Gagliardi, G. & Russo, I. (2012a), The IMAGACT Cross-linguistic Ontology of Action. A new infrastructure for natural language disambiguation. In N. Calzolari, K. Choukri, T. Declerck, M.U. Doğan, B. Maegaard, J. Mariani, J. Odiijk & S. Piperidis (eds). *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. European Language Resources Association (ELRA): PARIS, pp. 2606-2613.
- Moneglia, M., Gagliardi, G., Panunzi, A., Frontini, F., Russo, I. & Monachini, M. (2012b), IMAGACT: Deriving an Action Ontology from Spoken Corpora. Paper presented at the *Eighth Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-8)*. Pisa, October 3-5, 2012.