Translating Action Verbs using a Dictionary of Images: the IMAGACT Ontology

Alessandro Panunzi°, Irene De Felice*, Lorenzo Gregori°, Stefano Jacoviello†, Monica Monachini*, Massimo Moneglia°, Valeria Quochi*, Irene Russo*
°University of Florence, *ILC CNR (Pisa), †University of Siena
alessandro.panunzi@unifi.it, irene.defelice@ilc.cnr.it, lorenzo.gregori@unifi.it,
stefano.jacoviello@gmail.com, monica.monachini@ilc.cnr.it, moneglia@unifi.it,
valeria.quochi@ilc.cnr.it, irene.russo@ilc.cnr.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 in all translation 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 IMAGACT search interface and the various kinds of 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 descriptions.

Keywords: Action verbs; Image ontology; Multilingual dictionary; Computer-aided translation

1 Introduction

In all language modalities, action verbs bear the basic information that should be understood in order to make sense of a sentence. Moreover when we communicate, we have to refer to actions very often. Native 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 the more common action verbs, in their own meaning, refer to many different actions: in this sense, they are "general" verbs. Moreover, each language categorizes actions in its own way. These facts imply that there are not one-to-one translation relationships between different general verbs in different languages (Majid et al. 2007; Kopecka & Narasimhan 2012). If we take the English verb *to*

cross, for instance, we could argue that it can refer to at least two different action types, as in the sentences:
(1) John crosses the street
(2) John crosses his arms
On the contrary, in Italian we must use two different verbs to translate the previous sentences, namely *attraversare* (for *crossing the street*) and *incrociare* (for *crossing arms*):
(3) Gianni attraversa la strada
(4) Gianni incrocia le braccia

The problem is a significant one because reference to action is very frequent in ordinary spoken communication (Moneglia & Panunzi 2007) and specifically high-frequency verbs can each refer to many different action types (Moneglia in press).
The IMAGACT project has now delivered a corpus-based language ontology 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 and visually represented them by means of prototypical scenes, either animated (3D) or filmed (Moneglia et al. 2012; Frontini et al. 2012). The cross-linguistic correspondences to action concepts of 521 Italian verbs and 550 English verbs (i.e., the verbal lexicon most likely to be used when referring to action) are stored in a database. The action concepts in IMAGACT have already been extended to Chinese and Spanish (included in the first IMAGACT release). Perhaps more importantly, the action concepts can be easily identified by speakers of any language, since they are represented in an ontology of animated and filmed scenes.
This paper presents the IMAGACT online interface and how queries are made to the database. The user can search in IMAGACT in three main ways: a) as a bilingual dictionary, based on concept selection; b) through explicit comparison of the range of actions that can in principle be referred to by two lexical entries, of the same language or of different languages; c) through the direct selection of an action concept in the gallery of prototypic scenes, independently of the language of the user. In the last section, the paper also introduces an initiative aimed at the extension of the IMAGACT database to other languages.

2 Dictionary

If the user wonders how an English action verb translates into Italian or into another target language (Spanish and Chinese in the IMAGACT first release), IMAGACT can be used as a multilingual dictionary of images. Figure 1 shows the thumbnail images of the main types of actions identified by the English verb *to cross*.

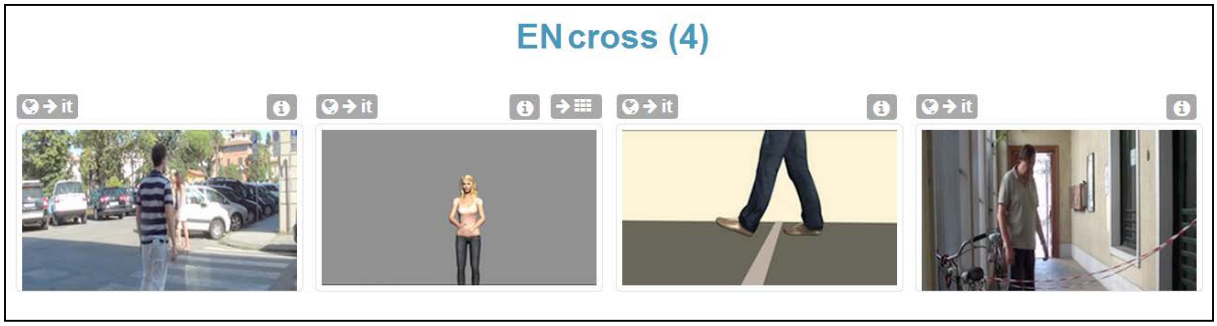

Figure 1: The variation of *to cross* across action types.

Looking at the various action types this verb expresses, the user can:

- select the action type he is interested in
- look at the animation to clarify the referred action
- see how this action is identified in the target language

IMAGACT returns one main verb and an additional set of verbs which equally identify this specific type of action. For each scene, which represents a distinct action type, Italian gives different translation for this verb, as shown in Figure 2.

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

3 Comparison

The user can compare verbs that in principle should translate between each other from two different languages. Searching with this function, the system illustrates the set of action types in which both verbs can be respectively applied. The result of such a search for *to cross* and *attraversare* (see Figure 3) supports the intuition that the two verbs can translate to each other, at least with respect to some of the action types they can refer to. 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 *attraversare* and the English verb *to cross* becomes explicit. The Italian user will learn that, in English, *to cross* cannot be applied to the types on the right column in Figure 3. In this case, he can go directly to the English translation of verb *attraversare*, as shown in Figure 4: for these two actions he has to use, respectively, *to traverse* / *to pass* and *to stab* / *to pierce*.

Comparison between two verbs can also be requested within the same language, to allow the user exploring more deeply the differences in meaning between the lexical entries suggested by the system. For instance, an English user can learn that both the verbs *passare* and *attraversare* can be applied to the action type illustrated by the scene on the left column, second row, in Figure 4. The user may wonder what the difference is between the two Italian lemmas suggested by the system. So, he can then compare the two verbs (of the same target language), clarifying the differences between their referential properties (Figure 5).

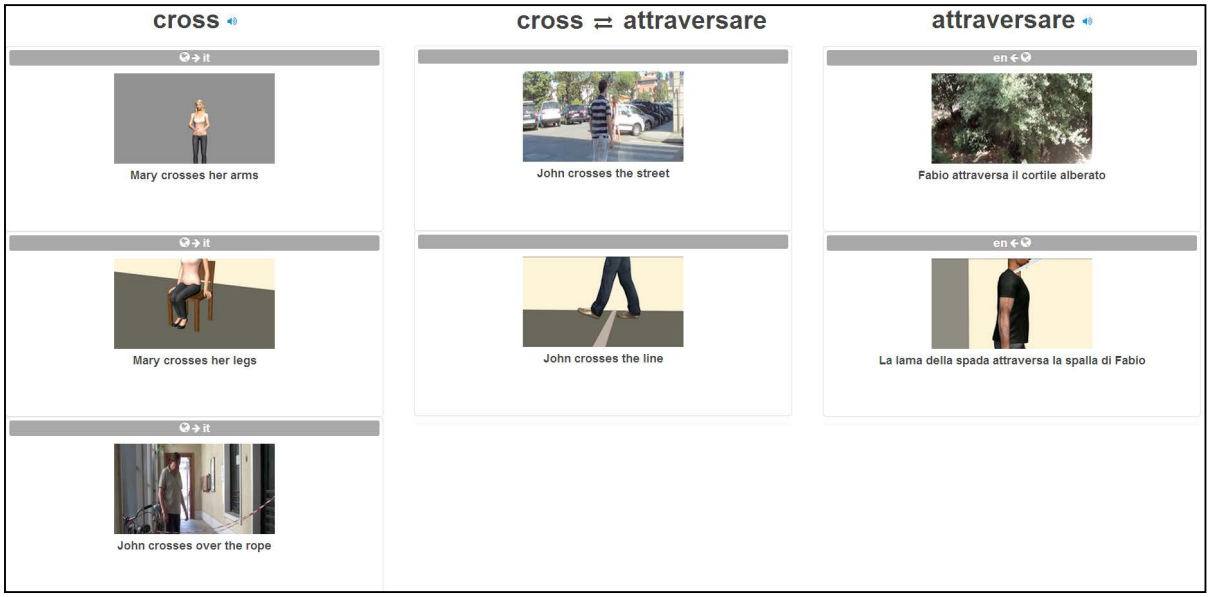

Figure 3: Comparison of *turn* vs. *girare* (results of the query interface with graphic adaptations).

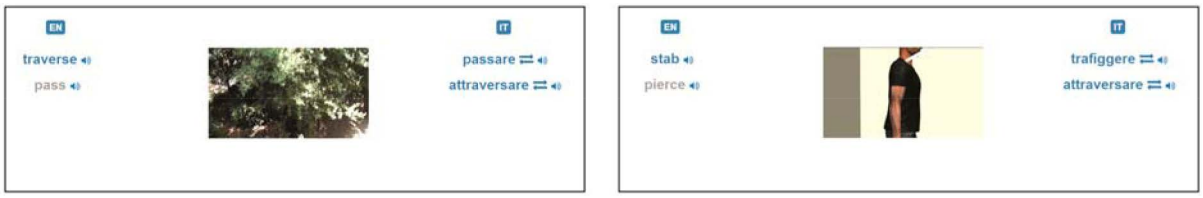

Figure 4: From comparison to linguistic categorization.

Figure 5: Intra-linguistic comparison.

4 Gallery

If the language of the user is not represented in IMAGACT, he 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 through the selection of one "meta-category" of action among the ones proposed by the interface. Such meta-categories are represented by a series of 3D animations, which are continuously played in loop, as the thumbnails in Figure 6 suggest.

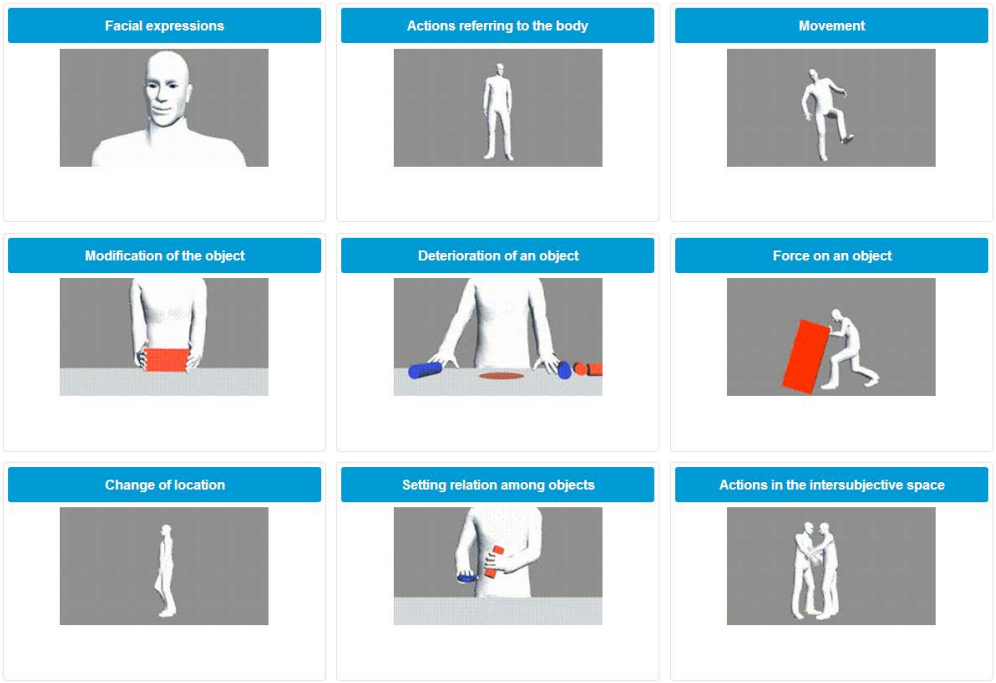

Figure 6: Representation of action meta-categories through avatars.

The numerous actions covered by IMAGACT are gathered into 9 macro-classes, which have high relevance in human categorization of action. Meta-categories are ordered according to criteria which take into account the informative focus of the action, as reported in Table 1.

Perspective centered on the Actor	Perspective centered on the Actor-Theme relation	Perspective centered on the Theme-Destination relation
Actions referring to facial expression	Modifications 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	Forces on the object	Actions in inter-subjective space

Table 1: Criteria for meta-categories.

The user can figure out what kind of action these stand for by looking at the abstract representation heading each class, and of course through a quick look at the actions gathered under each one. The user identifies the action he is interested in independently of the word he has for that action in his own language; after choosing the action via its visual representation, he is able to reach its linguistic categorization in the required target language. From this point of view, the IMAGACT gallery reverses the ordering of the dictionary: it goes from concepts to language instead of from language to concepts.

Once the user has understood the meaning of the action groups, it will be easier to search for the specific action he is interested in. He 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.

For instance, Figure 7 is what the system returns when asked for the Chinese verb for the action corresponding to the verb *to cross* under the category *Actions referring to the body* (i.e., *crossing the arms*).

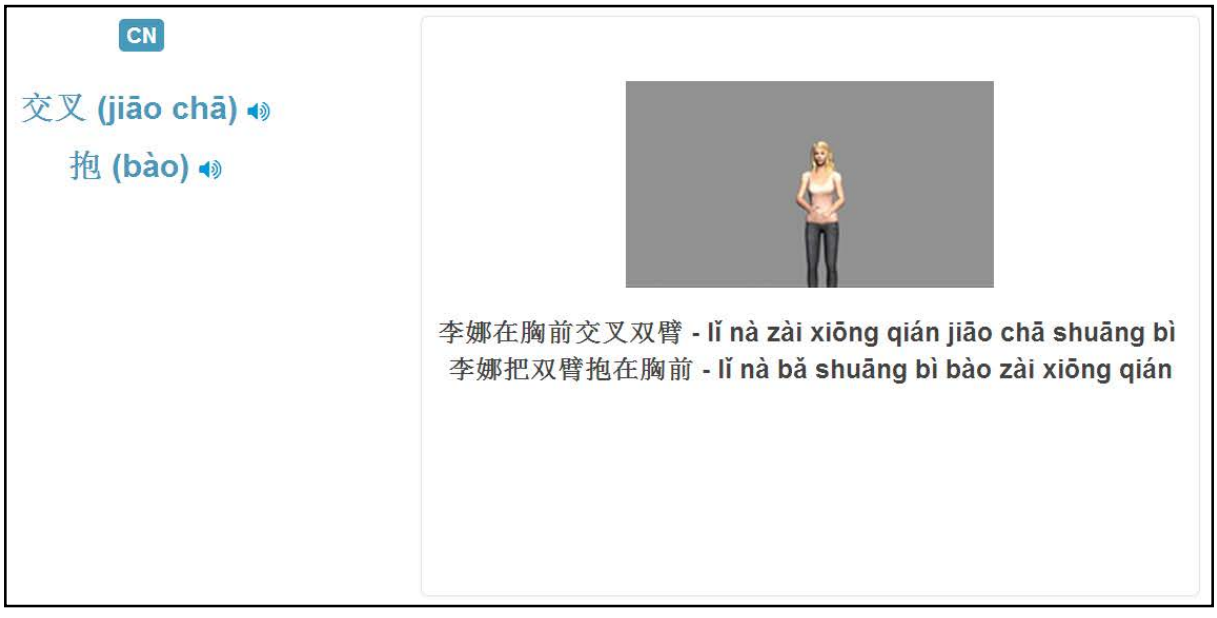

Figure 7: From gallery to linguistic categorization (Chinese).

5 Extending the dictionary

Because IMAGACT's direct representation of actions through scenes can be interpreted independently of language, the system 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 (IMAGACT4ALL).

In the simplified interface for the Competence Based Extension of the IMAGACT database to other languages (called *CBE light*), the set of action concepts represented by the IMAGACT prototypic scenes is assumed as a fixed-reference universe, and the work starts directly from such scenes.

An informant receives the action types as input. Figure 8 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 its citation 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, third-person singular, 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 Latin characters, the informant also provides the verb and its caption in Latin characters, as can be seen for Chinese.

Figure 8: Simplified Competence Based Extension (CBE light).

Given that verbs with different meanings can identify the same action, the informant is asked to find multiple lemmas allowed by his language for each action. However, simply viewing one short clip may be not sufficient to elicit all the 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 sugge-

stions 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 his language, translations that can be used in referring to the event in question. If so, he will report a new verbal lemma and a new caption by adding a line to his language options. The work of the informant must be supervised by a mother-tongue expert linguist before the language is mapped onto the IMAGACT database. 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 extend the database to a number of different languages.

6 References

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. In M. Zock, R. Rapp (eds) *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex III), Mumbay, 15 December 2012*. Red Hook (NY): Curran Associates, pp. 69-80.

IMAGACT. Accessed at: http://www.imagact.it [04/04/2014].

Kopecka, A., Narasimhan, B. (2012). *Events of Putting and Taking, A Cross-linguistic perspective*. Amsterdam/Philadelphia: John Benjamins.

Majid, A., Bowerman, M., van Staden, M., Boster, J. S. (2007). The semantic categories of cutting and breaking events: A crosslinguistic perspective. In *Cognitive Linguistics*, 18(2), pp. 133-152.

Moneglia, M. (in press). The semantic variation of action verbs in multilingual spontaneous speech corpora. To appear in T. Raso, H. Mello (eds) *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: John Benjamins.

Moneglia, M., Monachini, M., Calabrese, O., Panunzi, A., Frontini, F., Gagliardi, G., Russo, I. (2012). The IMAGACT Cross-linguistic Ontology of Action. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis (eds) *Proceedings of Eighth Language Resources and Evaluation Conference (LREC 2012), Istanbul, 23-25 May 2012*. Paris: ELRA, pp. 2606-2613.

Moneglia M., Panunzi A. 2007. Action Predicates and the Ontology of Action across Spoken Language Corpora. The Basic Issue of the SEMACT Project. In M. Alcántara Plá, Th. Declerk (eds) *Proceeding of the International Workshop on the Semantic Representation of Spoken Language, SRSL7, at Conferencia de la Asociación Española de Inteligencia Artificial, CAEPIA 2007, Salamanca, 12-16 November 2007*. Salamanca: Universidad de Salamanca, pp. 51-58.

Acknowledgements

The IMAGACT project is funded by the PAR/FAS program of the Tuscan region in Italy. Further research on the IMAGACT database, including this paper, is being accomplished with the contribution of the MODELACT project (2013-2016), funded by the national programme Futuro in Ricerca.