

IMAGACT - A Cross-linguistic Visual Ontology of Action

A system for learning how to refer to Actions in English, Italian, Spanish and Chinese

USER GUIDE

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.

Like second language learners, automatic translation systems also suffer from this problem, especially when the translation of a simple sentence is required.

For instance, asking the Google system to translate into Italian the English sentence

"John *pushes* the garbage into a can" we presently get "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" is translated into English as "Mario *takes* the cat by the tail"
instead of "Mario *catches* the cat by the tail", which is the actual interpretation;

"Mario *gira* gli zucchini nella padella" is returned as "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 ontology.

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

The IMAGACT project, funded by the PAR/FAS program of the Tuscan region in Italy, has now delivered a corpus-based 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 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, Italian, Spanish and Chinese have been established in a data base that can be freely accessed on line through an user friendly user interface.

The verbal lexicon most likely to be used when referring to action in the above languages is stored in the database. Each verb can express one or more concepts, while each concept can refer to one or more verbs (within and across languages).

This ontology gives a picture of the variety of activities that are prominent in our everyday life and specifies the verb used to express each one in ordinary communication.

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 guide presents the IMAGACT internet infrastructure and how queries are made to the database. For each of the above languages, the interface makes explicit the range of variation of each action verb across the action types. It also specifies the cross-linguistic correspondences that are possible for each represented action.

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.

Dictionary (multimedia in <http://www.youtube.com/watch?v=azi9hgXo9mQ>)

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 I can be used as a multilingual dictionary of images.

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, look at the animation to clarify the meaning, and 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. In almost all cases for the verb *to turn*, such as the scene with the postcard, he will see that the appropriate entry in Italian is the verb *girare*.

However, among the set of actions referred to by *to turn*, the user may be interested in the one highlighted, where *turn* is equivalent to *fold*. IMAGACT reveals that this action requires the use of a different verb in Italian (*ripiegare* or *rigirare*). Figure 2 shows how the system returns the information.

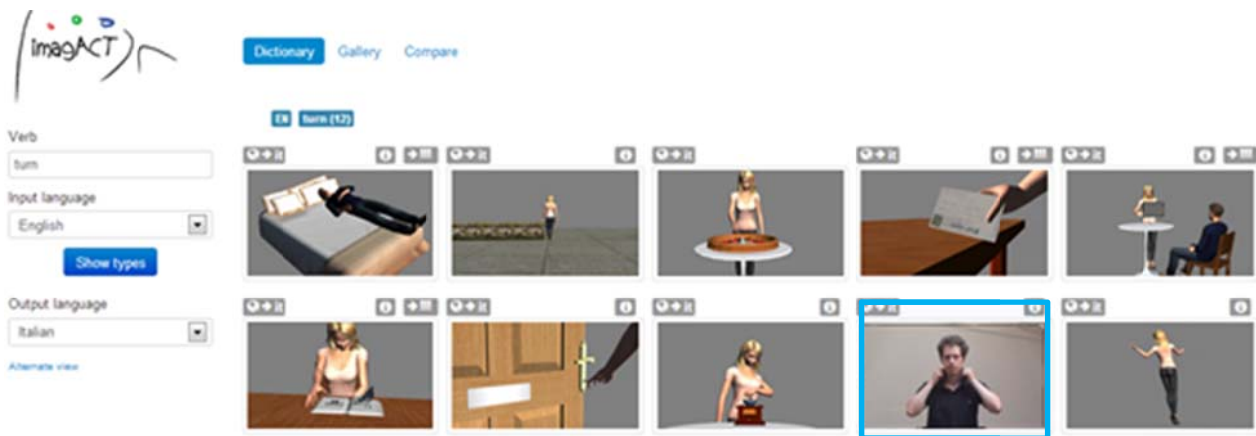


Figure 1 The variation of *to turn* across action types



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

Some of the types shows this icon on top on the right.

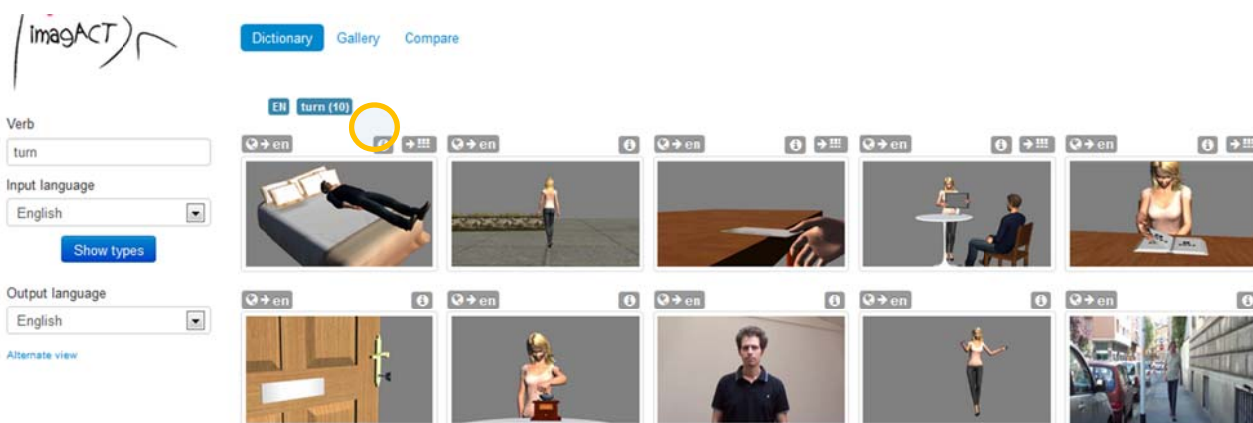


Figure 3 Access to *Families*

It means that the type is very general and that it gather a family of types . Types in a family are all very similar to the one presented, but each one bearing some difference that have linguistic relevance

For instance this family of types refers to rotations made with the body that can be all identified by *to turn* in English and *girare* in Italian



Figure 4 Type expanded in a family

However each type in the family contains features which, in both languages, require a different set of equivalent verbs



Figure 5

In this type to turn is equivalent to *flip* and *roll* and, besides the general verb *girare*, which applies in all cases, can be also identified in Italian by *rivoltare* and *rigirare*



Figure 6

This type is still a body rotation, but in English it is equivalent to *twist* instead, and *girare* also allows a totally different set of equivalent verbs

So, IMAGACT allow to know how languages refer to action in details. Besides the main variation of verbs accross actions, it also focus on subtle differences conveyed by action verbs in its languages

Comparison (multimedia in <http://www.youtube.com/watch?v=2nmZ0mhQQS4>)

IMAGACT can be used to learn the lexicon of a foreign language using a specific method of language acquisition. The user can compare 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* (Fig. 7) 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.

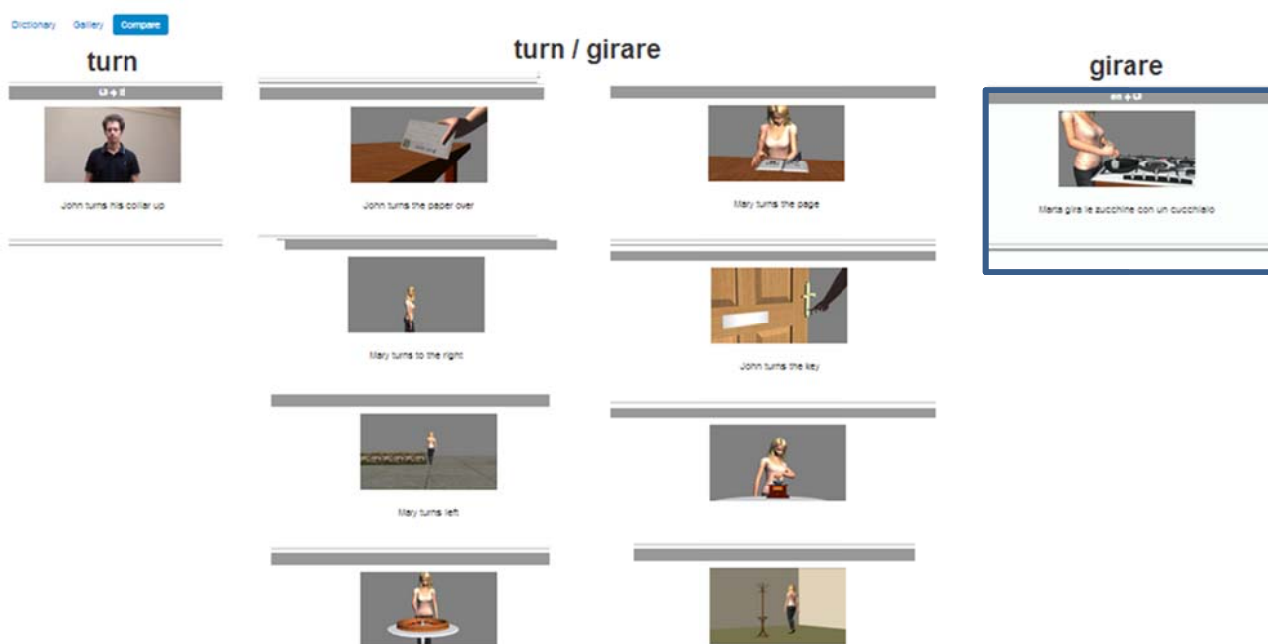


Figure 7. Comparison of *turn* vs. *girare* (results of the query interface, with graphic adaptations)

The Italian user will learn that in English *turn* cannot be applied to the type that is highlighted below. The verb *stir* should be used instead.

Comparison can be both across and within language.

Comparisons within language allow the user to explore more deeply the subtle differences in meaning between the lexical entries suggested by the system.

Each action type among those represented by a particular verb is also related to a set of equivalent verbs. This raises the question of what difference the choice of one verb or another conveys in the target language.

For instance, an English user focusing on the case in Figure 8 may wonder what the difference is between the two Italian lemmas suggested by the system (*girare* / *mescolare*).

EN
stir



IT
girare ⇌
mescolare ⇌

Marta gira le zucchine con un cucchiaino

Figure 8

To answer this question, the user can compare two verbs of the target language as in Figure 9








girare	girare / mescolare	mescolare
<p>it → it</p>  <p>Marta gira la manovella del macinacaffè</p>	 <p>Marta gira le zucchine con un cucchiaino</p>	<p>it ← it</p>  <p>Marta mescola uova e farina in un piatto</p>
<p>it → it</p>  <p>Marta gira pagina</p>		<p>it ← it</p>  <p>Marta mescola le carte</p>
<p>it → it</p>  <p>Marta si gira</p>		<p>it ← it</p>  <p>Marta mescola i due colori con il pennello</p>

Figure 9 Intra-linguistic comparison

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. Therefore, the user can see that using *girare* in this case would focus more on the agent's motion, whereas *mescolare* would focus more on the effect of the action on the themes.

Gallery (multimedia in <http://www.youtube.com/watch?v=ajfX979nW3c>)

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.

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

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.

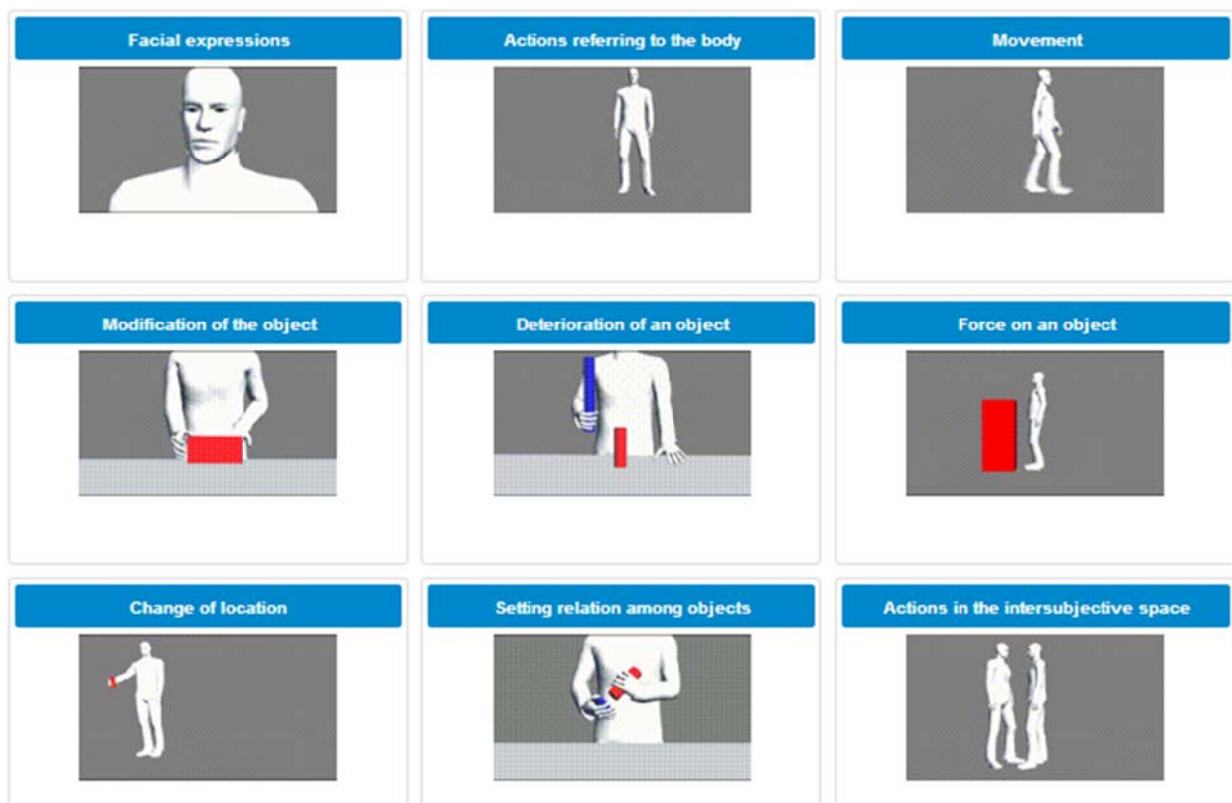


Figure 10 Representation of Action Classes through avatars

Once the user has understood the meaning of these 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.

This process is independent of the word she has for that action in her language. For instance, the following is what the system returns when asked for the Chinese verb for the action corresponding to *turn* under the category *Intersubjective space*.



Figure 11 From scene to linguistic categorization in Chinese

..... Have a nice trip in IMAGACT