Abstract

Action verbs are the least predictable linguistic type for bilingual dictionaries and they cause major problems for NLP technologies. This is not only because of language specific phraseology, but it is rather a consequence of the peculiar way each language categorizes events. In ordinary languages the most frequent action verbs are “general”, since they extend productively to actions belonging to different ontological types. Moreover, each language categorizes actions in its own way and therefore the cross-linguistic reference to everyday activities is puzzling. A cross-linguistic stable ontology of actions is difficult to achieve because our knowledge on the actual variation of verbs across types of actions is largely unknown. This paper briefly presents the problems and the building strategies of the IMAGACT Ontology, which aims at filling this gap, and compares some early results on a set of Italian verbs with the information contained in ItalWordNet.

1 The Semantic Variation of Action verbs within and across Languages

Action verbs refer to at least an eventuality where an agent/causer theta role fills its argument structure. In all language modalities, action verbs bear the basic information that should be processed in order to make sense of a sentence. Especially in speech, they are the most frequent structuring elements (Moneglia and Panunzi, 2007), but unfortunately no one-to-one correspondence can be established between an action verb, conceived as a lexical entry, and an action type, conceived as an ontological entity.

For instance, the English verb to take can refer to qualitatively different actions. In some uses the agent assumes the control of an object and changes its location (1); in some other uses the agent receives the object (2); in other cases, the agent takes the object away from somebody else (3):

1) John takes the umbrella
   [= to get]

2) John takes the book from Sara
   [= to receive]

3) the thief takes the money from the girl
   [= to take away]

In short, in the above circumstances different action types occur. This judgment is confirmed by the productivity of each action type. For instance, despite the fact that the predicate is applied to different objects, humans are able to judge by reading a set of sentences whether the same action is performed or not:

1a) John takes the glass/ the candle
(2a) John takes the pen from the assistant
(3a) the thief takes the hat off the lady

Moreover, to take has several meanings corresponding to the different action types, and none of these types can be considered more appropriate than the others in characterizing the meaning of the verb. Each one could be a prototypic instance of the verb (Givon, 1986).
We call general verbs all natural language action verbs that share this property. In the case of general verbs, ordinary language does not mirror the ontology of action, causing a huge problem for all natural language understanding and machine translation tasks. As a matter of fact, the lemma does not specify the referred ontological entity.

2 Action Ontology and Translation

The problem of the lack of one-to-one mapping between lexical entries and ontological entities becomes even more relevant when cross-linguistic communication is taken into account.

The above variation of to take is also shared by the verbs roughly translating it in Italian. However, no translation equivalent can be established between action predicates of the two languages, as far as the ontological entity referred by action verbs is not identified and there is no guarantee that two predicates in a bilingual dictionary pick up the same entity (Majid et al., 2008). For this reason, action verbs are puzzling for machine translation, which may not find the translation equivalent even for simple sentences.

For example, according to pragmatic circumstances, the Italian sentence in (4) can be interpreted as an instance of different actions and can be translated into English respectively with to take / to hold / to catch, but this can be properly foreseen only if action types are identified cross-linguistically:

(4) Mario prende il gatto
(4a) Mario takes the cat
(4b) Mario holds the cat
(4c) Mario catches the cat

Given that action verbs have high frequency both in speech and in written corpora, this problem is extremely important for practical applications. The existence of the above semantic relations cannot be predicted, since they require general ontological knowledge which is not accessible through lexical entries.

Nevertheless, the mapping of general verbs to the action types is productive and should be in principle predictable. Once one action type is identified, we can foresee that the translation relation among predicates referring to that type in different languages holds in all instances of the type.

Therefore, action types can be considered as an ontological level that is independent from the language.

2.1 Action ontology and lexical databases: primary and marked variation

Existing verb typologies have gone a long way in systematically categorizing verbs into classes.

There is a range of lexical resources and ontologies which provide information on verb meaning variation (Baker et al., 1998; Levin, 1993; Kipper-Schuler, 2005; Palmer et. al., 2005) and a number of initiatives which extend the information provided according to each frame for many languages.

The problems encountered by present ontologies to deal with categorization of action at cross-linguistic level can be made explicit by looking at WordNet (Fellbaum, 1998). Verbs are an important part of WordNet: more than 11,000 lexical entries, divided into 25,047 senses and corresponding to 13,767 synsets in the English database (version 3.0), which has been extended to many other languages.

For instance, WordNet identifies 42 synsets for the verb to take. Let’s just focus on three of these entries:

a) S: (v) take, get hold of (get into one's hands, take physically) "Take a cookie!"; "Can you take this bag, please";

b) S: (v) lead, take, direct, conduct, guide (take somebody somewhere) "We lead him to our chief"; "can you take me to the main entrance?"; "He conducted us to the palace";

c) 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".

Despite its richness, this information is hard to use for disambiguation and translation tasks even by expert annotators (Ng et al., 1999). Since the glosses given for each synset are often too vague, the identification of the actual use of a verb among all its synsets becomes difficult.

Another crucial reason is that the productivity of verb application cannot be guaranteed by all synsets in the same manner. More specifically WordNet does not distinguish the synsets instantiating the proper meaning of the verb (for in-
stance a and b) from those which instantiate phraseological or metaphorical usages (for instance c).

Verbs have various usages which depart from their actual meaning, but those usages do not constitute any productive action type. From this perspective, it is reasonable to foresee that the Italian verb prendere can be applied to all instances of a (5) and in no instances of b (6):

(5) he takes a cookie / a glass / a bag
(5a) lui prende un biscotto / un bicchiere / una borsa
(6) he takes the car / the dog / his friend there
(6a) *lui prende la macchina / il cane / il suo amico là

On the contrary this is not the case of c (7-8), which is a metaphorical usage of the verb. We cannot foresee any regularity in the application of the Italian verb prendere to the possible instances of c:

(7) he took an air of superiority
(7a) ha preso un’aria di superiorità
(8) he took on strange manners
(8a) *ha preso strane maniere

In summary, despite the high number of senses registered in WordNet, there is no possibility of identifying those types that constitute the basis for a productive cross-linguistic relation.

3 The IMAGACT project

The IMAGACT project, which has been funded in Italy with the PAR/FAS program (undertaken by the University of Florence, ILC-CNR, Pisa, and the University of Siena) uses both corpus-based and competence-based methodologies for simultaneous extraction of a language independent action inventory from spontaneous speech resources of different languages.

The IMAGACT infrastructure faces key issues in ontology building. It grounds productive translation relations since it distinguishes the proper usage of verbs from their metaphorical or phraseological extensions; it allows easy identification of types in the variation, it is cross-linguistic in nature, it derives from the actual use of language but it can be freely extended to other languages through competence-based judgments and it is therefore suitable for filling gaps in lexical resources.

3.1 Exploiting spontaneous speech corpora

The first idea developed in IMAGACT is to strictly define the relevant domain of language usage from which data about linguistic reference to actions can be derived. Actions specified by those verbs are most frequently used in ordinary communication since they are very relevant in everyday life. The actual use of action oriented verbs in linguistic performance can therefore be appreciated by observing their occurrence in spontaneous speech corpora, in which reference to action performance is primary.

The IMAGACT database focuses on high frequency verbs, which can provide sufficient variation in spoken corpora; i.e. roughly 500 verbs referring to actions which represent the full basic action oriented verbal lexicon.

In order to maximize the probability of occurrence of relevant action types, IMAGACT identifies the variation of this set in parallel on two spoken corpora:

- a 2 million word English corpus, taken from the British National Corpus;
- a collection of spoken Italian corpora with 1.6 million words in total (LABLITA corpus, Cresti and Moneglia, 2005; LIP, De Mauro et al., 1993; CLIPS corpus).

The corpus-based strategy consists of a manual annotation of the instances of action verbs, which first separates the metaphorical and phraseological usages from proper occurrences and then classifies proper occurrences into action types.

3.2 The cross-linguistic definition of the ontology of action in a Wittgenstein-like scenario

The experience in ontology building has shown that the level of consensus that can be reached in defining entities which are object of language reference tends to be lower, since the identification of such entities relies on a definition. Definitions are highly underdetermined, since they depend on the granularity of feature retrieval.

The traditional methodology will require reconciling in a unique definition all definitions given by linguists to classify the actions occurring in each language corpus. This is certainly a difficult task.
The key innovation of IMAGACT is to provide an alternative methodology which exploits the language independent ability to recognize similarities among scenes, distinguishing the identification of action types from their lexicographic definition. The annotator is required to identify the action that is expressed in each instance of the action verbs taken into account. Different actions are identified for each verb and then grouped into cross verbal action types.

All primary occurrences (the ones referring to a physical action) of a verb in the corpus are manually clustered around best examples that represent the different actions in terms of involved body schema, spatial relations and focal properties. Local equivalents (other verbs that can be replaced in the instance) are identified for each cluster, thus producing verb independent action types. Action types are then described in scripts and videos are made, onto which action types from many languages are mapped.

Working with more than one language produces a language independent type inventory. Crucially only the identification (and not the active writing of a definition) is required to set up the cross-linguistic relations. In Wittgenstein’s terms, how can you explain to somebody what a play is? Just point out a play and say “this and similar things are plays” (Wittgenstein, 1953).

In IMAGACT the ontology makes use of the universal language of images which allows reconciling in a unique inventory of action types the descriptions derived from the annotation of corpora belonging to different languages.

For instance, let us consider the Italian verb *spingere* and the English verb *to push*, which might be expected to match on a similar set of action types. The annotation of the Italian corpus has identified six different types extended by *spingere*, which are instantiated by the following best examples:

9) il dottore spinge sulla pancia del paziente
[the doctor presses on the patient’s belly]
10) John spinge la carta nel cestino
[John pushes the paper in the basket down]
11) la ragazza spinge il carrello
[the girl pushes the trolley]
12) Maria spinge la bottiglia giù dal tavolo
[Mary pushes the bottle off the table]
13) la ragazza spinge la creta nello stampo
[the girl pushes the plasticine into the mold]
14) lo yogi spinge il ventre avanti
[the yogi pushes his belly out]

On the basis of this information, a scene representing the occurrence of each type is produced. Therefore the above best examples will be respectively linked to A1, A5, A6, A8 and A9 of Figure 1 below.

Assuming that the English corpus will also be processed, the action types extended by to push will come about through best examples extracted from the corpus, obviously referring to different eventualities.

Of course there is no necessity that all possible types extended by *to push* and *spingere* will be recorded in the corpora, however the intersection of types actually extended by both verbs can be easily recognized. Indeed on the basis of the evidence provided by the cited scenes all competent speakers will recognize that, for instance, the Best examples “Mary pushes the car on the highway” should be mapped onto A5 and “The killer pushes the man off the cliff” should be mapped onto A6. This will be achieved without any direct comparison between Italian and English.

On the basis of the scenes, the differential of the two verbs for what regards their possible extension across action types recorded in the corpus will also easily be recognized. For instance, competent speakers will also recognize that A7 is the only model of Figure 1 in which to push cannot be extended at all. Considering the difference in extension between *spingere* and to push it will become evident that to push is more general since it can also be extended to A3, which is not a possible model of *spingere*.

IMAGACT will deliver a database of action types with their language encoding of English and Italian verbs in conjunction with the set of sentences (derived from corpora) instantiating each type.

Action types will be recorded in the form of videos. The scene corresponding to the best example of each action type (prototypic scene) is played by a supervisor. The adequacy of this scene in representing what’s specified in the annotation is negotiated by the supervisor with the annotator to avoid misunderstandings. The scene is then transformed into a 3D animation and all information that is not essential to the representation is eliminated (stereotypic scene, not available at this stage of the project).

On the basis of this outcome it will be possible to ask informants with a different language what verb(s) is applied in their language for each type
identified by a scene and by a set of English sentences derived from corpus occurrences and assigned to that scene. The informant will provide the lexical choice available in his language. Crucially, the informant will verify whether or not the choice is correct for all arguments retrieved from corpus and assigned to that type.

4 Linking to WordNet

The IMAGACT project has already produced a corpus-based extraction of action types from a subset of high frequency Italian action verbs.

Let’s consider as an example a restricted set of Italian verbs roughly equivalent in some way or other to *to press* or *to push*:\{spingere, premere, schiacciare, pigiare, comprimere, spremere, pressare\}.

The IMAGACT methodology starts from the identification of the semantic (referential) variation of the verbs. Once this variation is identified, it is possible to list all the referred action types, and then to connect to each type the set of equivalent verbs that can be applied to it.

From the given set of verbs a tentative inventory of action types has been extracted (see Figure 1).

Of course, each of the verbs that have been taken into account can be used to express a subset of these actions. For instance the verb *spingere* (the more general in the examined cases) can be applied to the actions A1, A2, A3, A5, A6, A8 and A9, while the verb *spremere* can be applied only to action A3.

Conversely, all identified action types can be referred by a set of verbs: Action A1 can be expressed by *spingere*, *premere*, *pigiare*, while Action A3 can be expressed by *schiacciare*, *comprimere*, *spremere*.

<table>
<thead>
<tr>
<th>ACTION TYPE</th>
<th>DEFINITION (and example)</th>
<th>EQUIVALENT VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Continuing generic pressure, with the sole result that the object (or body part) is pressed <em>(the doctor palpates the abdomen)</em></td>
<td>premere, pigiare, spingere</td>
</tr>
<tr>
<td>A2</td>
<td>Pressure on only one side that brings about the reduction of the volume of the object <em>(Jane compresses the garbage)</em></td>
<td>premere, pigiare, schiacciare, pressare, comprimere, spingere</td>
</tr>
<tr>
<td>A3</td>
<td>Pressure on two or more sides that brings about a reduction of the volume of the object <em>(John squeezes the toothpaste tube)</em></td>
<td>schiacciare, spremere, comprimere</td>
</tr>
<tr>
<td>A4</td>
<td>Non continuos (brief) pressure <em>(Jane presses the button)</em></td>
<td>premere, pigiare, schiacciare, spingere</td>
</tr>
<tr>
<td>A5</td>
<td>Continuous pressure that accompanies the object in the transition <em>(Jane pushes the trolley)</em></td>
<td>spingere</td>
</tr>
<tr>
<td>A6</td>
<td>Impulse that distances the object from the agent <em>(Jane pushes away the bottle)</em></td>
<td>spingere</td>
</tr>
<tr>
<td>A7</td>
<td>Pressure that causes damage to the object <em>(John crushes the tomato)</em></td>
<td>schiacciare</td>
</tr>
<tr>
<td>A8</td>
<td>Pressure that inserts the object into something <em>(John pushes the plasticine into the mould)</em></td>
<td>premere, pigiare, spingere</td>
</tr>
<tr>
<td>A9</td>
<td>Internal pressure <em>(the Yogi pushes the stomach out)</em></td>
<td>spingere</td>
</tr>
</tbody>
</table>

Figure 1: IMAGACT action types
Speculatively we expect to find synsets that match these groupings, thus in this case one corresponding to A1 \{spingere, premere, pigiare\}, one corresponding to A3, and so on.¹

Still in the perspective of a linking of the final IMAGACT action inventory to one or more WordNets, among which ItalWordNet (IWN, Roventini et al., 2003), we could imagine at least three kinds of links:

- **Perfect matching**: an IMAGACT type of action matches a synset;
- **IMAGACT action types enriches IWN**: one or more IMAGACT types of actions are subsumed by a synset;
- **IWN enriches IMAGACT action types**: an IMAGACT action type subsumes one or more synsets.

The possibility of an imperfect match can also be foreseen.

In order to carry out the linking, a set of basic heuristics have been defined and applied to better align corpus-induced action verb types and the IMAGACT action types with the lexical knowledge encoded in ItalWordNet.²

We don’t expect a full alignment but as a first step we aim at maximizing corpus-induced generalizations with synsets. In the early stages of the project we want to make clear the gap between lexical entries and ontological types.

<table>
<thead>
<tr>
<th>Act.</th>
<th>Synset</th>
<th>Type of match</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>(pigiare [1], premere [1], spingere [5]) “apply pressure to something”</td>
<td>hypernymic match (the synset captures the meaning of the action but is more generic, applies also to A4)</td>
</tr>
<tr>
<td>A2</td>
<td>(appiattire [1], schiaciare [1]) “flatten”</td>
<td>possible exact match</td>
</tr>
<tr>
<td>A2</td>
<td>(comprimere [1], pressare [2], schiaciare [2]) “compress”</td>
<td>hypernymic match (A3, A8)</td>
</tr>
<tr>
<td>A3</td>
<td>(comprimere [1], pressare [2], schiaciare [2]) “compress”</td>
<td>hypernymic match (A2, A8)</td>
</tr>
<tr>
<td>A3</td>
<td>(spremere [1]) “squeeze”</td>
<td>hyponymous match (applies only when the object is soft)</td>
</tr>
<tr>
<td>A4</td>
<td>(pigiare [1], premere [1], spingere [5]) “apply pressure to something”</td>
<td>hypernymic match (also A1)</td>
</tr>
<tr>
<td>A5</td>
<td>(imprimere un movimento [1], spingere [2]) “make something move, push”</td>
<td>possible exact match</td>
</tr>
<tr>
<td>A6</td>
<td>(spingere [1]) “push away”</td>
<td>Possible exact match</td>
</tr>
<tr>
<td>A7</td>
<td>(schiacciare [3], spiacciare [1]) “crush”</td>
<td>possible exact match</td>
</tr>
<tr>
<td>A8</td>
<td>(comprimere [1], pressare [2], schiaciare [2])</td>
<td>hypernymic match (A2, A3)</td>
</tr>
<tr>
<td>A8</td>
<td>(pressare [1]) “press”</td>
<td>no match</td>
</tr>
<tr>
<td>A9</td>
<td>(premere [3])</td>
<td>no match</td>
</tr>
</tbody>
</table>

¹ Let it be stressed that the action inventory that we are describing here is still under development, and it still lacks the important contribution that may derive from the analysis of English action verbs.

² The ItalWordNet lexical database (henceforth IWN) was first developed in the framework of the EuroWordNet project and then enlarged and improved in the national project SI-TAL1. The theoretical model underlying this lexicon is based on the EuroWordNet lexical model which is, in its turn, inspired by the Princeton WordNet (Fellbaum, 1998).
Table 1: IMAGACT to ItalWordNet linking
This table shows how the simple access to IWN via the lemmas of potential action verbs does not guarantee finding good matches for the action at hand. As foreseen, on the one hand we have some synset like (pigiare [1], premere [1], spingere [3]) and (comprimere [1], pressare [2], schiacciare [2]) that are less specific than the action. On the other hand, a synset like (spremere [1]) seems to be more specific in that it applies to a subset of the objects that are involved in A3. Finally, some lemmas point to IWN synsets that, after manual inspection, prove to be non-relevant for the action at hand because of the existence of marked/metaphorical meanings.

Some heuristics can be applied to fine tune the linking: one possibility would be to use words that the annotator has identified as local equivalents in order to try and disambiguate between synsets. Local equivalents are manually identified verbs that cover the same action type (a notion akin to but not equal to that of synonymy).

For example, uses like premere for the action type A2 (“Fabio preme la carta nel cestino / Fabio presses the paper in the bin”) are not directly represented in IWN but they are recoverable through the local equivalent verb pressare and schiacciare that have synsets matching that action type. This strategy is useful but not resolute, because in several cases it is not possible to recover missing synsets or to narrow too general synsets, simply by looking at local equivalents’ synsets. However, in this case study we found that this heuristic was effective for 2 IMAGACT action types (A2, A3) out of 9.

The possibility of an automatic alignment of IWN with the action types in IMAGACT can also be taken into account. A possible strategy could be creating a link between each action and each synset showing a certain amount of match with the set of verbs expressing that action.

A perfect match would be that each of the verbs related to the action appear in the synset. This being rarely the case, an algorithm could compute similarity. When the same synset is linked to more than one action the link could be automatically identified as a hyponymic link, that is, IWN has fewer distinctions than the IMAGACT ontology. Clearly though this strategy is not error free: it does not perform any check on whether the synset is referring to a phraseological usage and most crucially it does not work when there is just one verb associated with an action, and that verb appears in many synsets. Moreover, the possibility still exist that a synset including a totally different set of verbs (non-generic verbs for instance) can be found in IWN that matches the action we want to link.

5 Conclusions and future work
In this paper we briefly show that action verbs are the less predictable linguistic type for bilingual dictionaries and they cause major problems for NLP technologies because no one-to-one correspondence can be established between an action verb and an action type.

The need for general ontological knowledge which is not accessible through lexical entries motivates the IMAGACT project. It will use both corpus-based and competence-based methodologies for simultaneous extraction of a language independent action ontology from spontaneous speech corpora for different languages.

Although the project was just started, several issues concerning the initial version of a stable ontology of actions are already evident from the case study presented in this paper. We didn’t expect a full alignment but we aim at maximizing as a first step corpus-induced generalizations with synsets. As a first result we have shown that actions that are different for human annotators are not always mirrored by equivalent entries in lexical resources.

The comparison between corpus-induced generalizations about action types and lexical information found in ItalWordNet gives rise to a set of heuristics (i.e. using the hierarchy of IWN, checking for local equivalents’ synsets) that can be useful in the near future for a cross-linguistic integration of action types performed on the basis of English WordNet, after which the annotation process for English will be finished.

Acknowledgments
The IMAGACT project is funded within the PAR FAS Program of Tuscan Region (Italy), Action 1.1.a.3.

References

CLIPS Corpus. URL: http://www.clips.unina.it


LABLITA Corpus of Spontaneous Spoken Italian. URL: http://lablita.dit.unifi.it/corpora/


