International Conference "ICT for Language Learning" 6th edition

IMAGACT E-learning Platform for Basic Action Types

Massimo Moneglia[1], Alessandro Panunzi[1], Gloria Gagliardi[1], Monica Monachini[2], Irene Russo[2], Irene De Felice[2], Fahad Khan[2], Francesca Frontini[2]
[1]Università di Firenze, [2]ILC CNR (Italy)
moneglia@unifi.it

Abstract
Action verbs express important information in a sentence and they are the most frequent elements in speech, but they are also one of the most difficult part of the lexicon to learn for L2 language learners, because languages segment these concepts in very different ways. The two sentences "Mary folds her shirt" and "Mary folds her arms" refer to two completely different types of action, as becomes evident when they are translated into another language (e.g., in Italian they would be translated as "Maria piega la camicia" and "Maria incrocia le braccia" respectively). IMAGACT e-learning platform aims to make these differences evident by creating a cross-linguistic ontology of action types, whose nodes consist of 3D scenes, each of which relates to one action type. In order to identify these types, contexts of use have been extracted from English and Italian spontaneous speech corpora for around 600 high frequency action verbs (for each language). All instances that refer to similar events (e.g., fold the shirt/ the blanket) are grouped under one single action type: each one of these types is then represented by a linguistic best example and a short video that represents simple actions (e.g. a man taking a glass from a table).The action types extracted for Italian and English are compared and merged into one cross-linguistic ontology of action. IMAGACT has provided an internet based annotation infrastructure to derive this information from corpora. The project is now completed for the Italian and English lexicon, data extraction for Chinese and Spanish is ongoing. Reference to prototypical imagery is crucial in order to bootstrap the learning process. By selecting the set of 3D scenes referred to by a verb in one language and viewing the type of activity represented therein learners can directly understand the range of applicability of each verb. Thanks to an easy interface, a user can access the English/Italian/Chinese lexicon by lemma or directly by 3D scenes. For example, searching for the verb "to turn",s/he will be presented with a number of scenes, showing the various action types associated to that verb.Clicking on a scene s/he or she will know how this type of action is referred to in other the languages

1. The IMAGACT project
The IMAGACT project, which has been funded in Italy with the PAR/FAS program of the Tuscan Region (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 framework 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 [1]. The result of the annotation work is accessed through the IMAGACT query interface, which is designed to be used to master the verbal lexicon of a second language during the acquisition process.
The project focuses on high frequency action verbs (approximately 600 lexical entries) of both Italian and English, which represent the basic verbal lexicon of action in the two languages. The purpose of the project is two-fold: (i) to derive information about action verbs adopting a bottom-up approach, from spoken corpora e.g the BNC [2] and C-ORAL-ROM [3]; (ii) to construct a multilingual ontology of action, anchored to videos. English and Italian action types derived from corpus are mapped onto the same set of scenes. The validation of basic action types so derived is accomplished through a procedure which maps the IMAGACT ontology onto an open set of languages. Chinese and Spanish are presently implemented. It is expected that the extension to more languages will make the basic action types more general and less dependent on a specific language, providing an innovative access to cognitively grounded information for language learning and teaching.

2. Annotation Infrastructure

The extraction of data from corpora is accomplished through a web based infrastructure [4] which is structured into two main interfaces, *Standardization & clustering of occurrences* and *Type annotation & assessment*. The tasks are accomplished by annotators with the assistance of a supervisor. The first task mainly aims at eliciting and describing the array of pragmatic situations that each action verb can extensionally denote. Action verbs carry the central information elements in a sentence and they are the most frequent items in speech [5]. Humans adopt the same verbal form to denote different types of events, as emerges from synonymic choices: for example, the verb "to take" in (1) *John takes a present from a stranger* means "to receive, to accept"; but in (2) *John takes Mary the book* it means "to bring"; in (3) *John takes the pot by the handle* it simply means "to grasp"; finally, in (4) *John takes Mary to the station* it means "to conduct". Every language shows a different behaviour in segmenting human experience into its action verbal lexicon. However we expect that, in a given language, similar events will be referred to by using the same verb: so "to take" will apply also to *John takes the children to school* / *his wife to the cinema*, similar to (4); we also expect this consistency to be found in other languages. These coherent clusters of similar events, denotable by the same set of action verbs, are referred to as action types. This kind of information was derived from around 100,000 spontaneous spoken language contexts found in the corpora. The procedure is the following:

- each occurrence of an action verb is extracted from English and Italian spoken corpora;
- linguistic contexts of each occurrence are then standardized and reduced to simple sentences (3rd singular form, present tense, active voice);
- proper instances, in which action verbs are used referring to actions, are distinguished from non-proper instances, in which action verbs do not refer to concrete actions or are used metaphorically);
- proper occurrences are grouped into action types, keeping granularity to its minimal level, so that each type contains a number of instances referring to similar events (*John takes the glass/the umbrella/the pen* etc.);
- from all standardized sentences of each type, one best example is chosen (or more than one, if the verb has more than one possible syntactic structure).

The annotator can read the larger context of the verbal occurrence and represents the referred action with a simple sentence in a standard form for easy processing. Along with the standardization, the annotator assigns each occurrence to a "variation class" thus determining whether or not it conveys the verb's meaning. This is what we mean by a PRIMARY occurrence. This task is accomplished through a synthetic judgment which exploits the semantic competence of the annotator [6] and is given in conjunction with Wittgenstein's hypothesis on how word extensions can be learned [7]. The occurrence is judged PRIMARY if: a) it refers to a physical action; b) it can be presented to somebody who does not know the meaning of the verb V, by asserting that "the referred action and similar events are what we intend with V". The occurrence is judged MARKED otherwise, as with "John rolls the words in his mind". Only occurrences assigned to the PRIMARY variation class make up the set of Action Types stored in the ontology. They must be clustered into families which constitute the productive variation of the verb predicate. The workflow thus requires the examination of the full set of standardized primary occurrences recorded in the corpus, whose meaning is now clear.
The infrastructure is designed to allow the annotator to create types ensuring both cognitive similarity among their events and pragmatic differences between them. Instances are assigned to the same type as long as they fit with one "best example". Clustered sentences should be similar as regards:

- the possibility to extend the occurrence by way of similarity with the virtual image provided by the best example (Cognitive Constraint);
- "equivalent verbs applied in their proper meaning" i.e. the synset [8](Linguistic Constraints);
- involved Action schema.

Among the occurrences the annotator chooses the most representative as best examples of the recorded variation, creates types headed by one (or more) best example(s), and assigns each individual standardization to a type by dragging and dropping.
Once all occurrences have been processed, negotiation with a supervisor leads to a consensus on the minimal granularity of the action types extended by the verb in its corpus occurrences. The relation to images of prototypical scenes provides a challenging question in restricting granularity to a minimal

family resemblance set: "can you specify the action referred to by one type as something like the best example of another?". Granularity is kept when this is not reasonable.
Once types are verified the infrastructure presents the annotator with the *Types Annotation & Assessment* interface. Conversely, in this task the annotator assesses that all instances gathered within each type can indeed be extensions of its best example(s), thus validating its consistency.
The assessment runs in parallel with the annotation of the main linguistic features of a type. The thematic roles are annotated using a restricted set of labels derived from current practices in computational lexicons (Palmer's Tagset in VerbNet) with adaptations [9]). When all proper occurrences of a verb have been assessed the annotator produces a "script" for each type and delivers the verb annotation to the supervisor for cross-linguistic mapping.

3. Cross-linguistic Ontology and Validation: from words to videos

Distinguishing families of usages of general verbs from the granular variations allows us to discover productive cross-linguistic relations, thus validating the ontology. This task aims at deriving from the data extracted a language-independent ontology for action.

- the two classifications of action types, derived independently from Italian and English and represented by best examples, are compared and merged into the same ontology of action types. Each node of the ontology is represented by a video exemplifying the whole action type;
- more than one Italian or English action type can be linked to the same video, when the verbs are locally equivalents (as for *taking something from someone* / *receiving something from someone*).

Once types of actions referred to by action verbs have been identified and the scripts have been produced for the best examples, with cross-linguistic equivalences established, the supervisor produces a prototypical scene. Actors perform the action described in the script or an equivalent action. The scene is recorded according to the following requirements, which are intended to reduce ambiguity and to trigger the preferred interpretation:

- Use of real-world objects instead of abstract/generic forms; minimal background information
- The scene is produced as an uninterrupted shot ("long take"); the action is performed with its usual temporal span (no slow-motion). The sequence is edited to focus on the relevant nucleus of the performed action (3-7 seconds)

The semiotic relevance of each scene and its capacity to elicit the appropriate verb is scrutinized by more than three experts before storage in the database. Subsequently a 3D animation is created from the videos, in order to make the scene even less ambiguous. The animation software used for the production of 3D videos is Autodesk MAYA3.
The result of the procedure described above is a set of short videos, each one corresponding to an action type, representing simple actions (see Fig. 2.).
Scenes represent the variation of all action verbs considered and constitute the IMAGACT ontology of action. This ontology not only is inherently interlinguistic, because it is derived through an inductive process from corpora of different languages, but also takes into account the intra-linguistic and inter-linguistic variation that characterizes action verbs in human languages.
To obtain a parallel corpus, all English standardized instances assigned to a type have been translated into Italian, and vice versa; the possibility of translating all instances of a type into another language, using only one verb, assures the coherence of that type. The validation of the ontology so derived is performed in parallel on English and Italian sentences gathered within each entry and generates a data set of parallel sentences.
A competence based extension to Chinese Mandarin is in progress and consists in identifying a verb in the target language for each type of the source language and verifying the applicability to all instances in the target language. Fig. 1 shows how this task is accomplished in the case of the second and fourth type of Fig. 2. The Chinese verbs *zhuàn* e *fàn* respectively work fine in all instances of the Italian verb *girare* in those types, showing that the application of the verbs is productive. In principle this procedure can allow the implementation of whatever language in the IMAGACT infrastructure.

Fig. 1. Validation & Extension interface

4. Conclusions: The IMAGACT e-learning platform

Fig.2a shows how the IMAGACT infrastructure returns the information regarding the variation allowed by the English verb *to turn*. This verb records 12 action types, most of which also are appropriate for the Italian verb *girare*. The English user learning Italian, however, can discover that the type highlighted does not allow this translation since, on the contrary it is not a possible type for the Italian verb (*il ragazzo alza* / *tira su il colletto* vs * *il ragazzo gira il colletto*). This information can be accessed by clicking on the icon, once Italian is chosen as output language.

If the user wants to go through an explicit process of language learning and is interested in mastering the difference between *to turn* and *girare*, IMAGACT provides him or her with an explicit method. The infrastructure allows one to compare the possible variations o*f to turn* with those of *girare*. Through the comparison interface the user will discover both the range of variations allowed by each predicate and their differential. Figure 2.b shows that, among the large set of intersection (only partially reproduced) the semantic competence underling the use of the verbs *girare* and *to turn* presents differences in very specific types. The representation of action through scenes is crucial. The learning process bootstrapped by pointing to prototypic 3D scenes allows the learner to foresee that the verb can be applied in all instances of the Action type i.e. it gives rise to the mastering of a productive concept.

This information can be acquired by a second language learner through natural language acquisition, however this requires a long exposure to language data. Moreover the productivity of types, which is evident to a mother tongue speaker, can never be taken for granted in a second language speaker. On the contrary the method provided in IMAGACT ensures an easy way to figure out how similar verbs of different languages can be used in different manners .

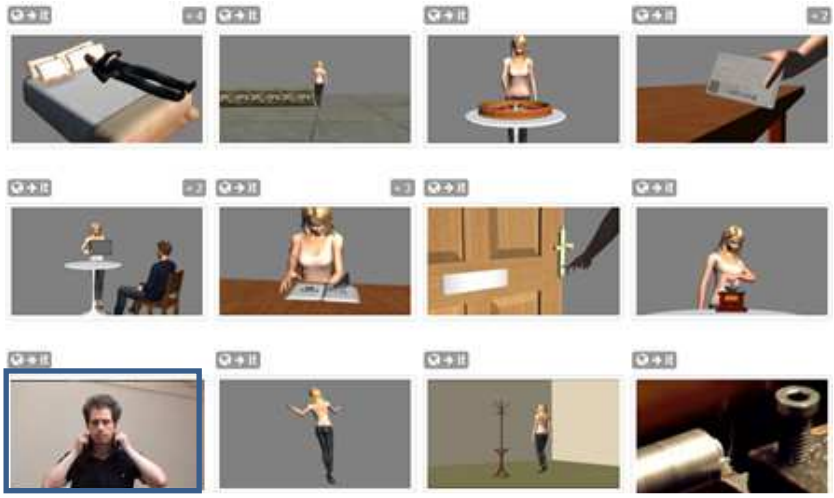

Fig. 2a. The variation of *to turn*

Fig. 2b. The comparison *to turn* and *girare*

References

[1] Moneglia, M. (2011). Natural Language Ontology of Action. A gap with huge consequences for Natural Language Understanding and Machine Translation. In Z. Vetulani (Ed.) Proceedings of the 5th Language & Technology Conference. Poznań: Fundacja Uniwersytetuim. A. Mickiewicza, pp. 95-100. [2] http://www.natcorp.ox.ac.uk/
[3] http://catalog.elra.info/product_info.php?products_id=757
[4] http://www.imagact.it/
[5] Moneglia, M., Panunzi, A. (2007). Action Predicates and the Ontology of Action across Spoken Language Corpora. In M. Alcántara, T. Declerck, Semantic Representation of Spoken Language (SRSL7). Salamanca: Universidad de Salamanca, pp.51-58.
[6] Cresswell M. F. (1978). Semantic Competence in F. Guenthner, M. Guenthner-Reutter, Meaning and translation. NY University Press: New York, 9-28
[7] Wittgenstein, L. (1953). Philosophical Investigations. Oxford: Blackwell.
[8] Fellbaum, Ch. (Ed.) (1998). WordNet: An Electronic Lexical Database. Cambridge: MIT Press.
[9]Palmer, M., Gildea, D. and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. In: Computational Linguistics, 31 (1), pp. 71-106.