IMAGACT: Deriving an Action Ontology from Spoken Corpora

Massimo Moneglia
Gloria Gagliardi
Alessandro Panunzi
LABLITA, University of Florence
moneglia@unifi.it

Francesca Frontini
Irene Russo
Monica Monachini
ILC, CNR, Pisa
monica.monachini@ilc.cnr.it

Abstract

This paper presents the IMAGACT annotation infrastructure which uses both corpus-based and competence-based methods for the simultaneous extraction of a language independent Action ontology from English and Italian spontaneous speech corpora. The infrastructure relies on an innovative methodology based on images of prototypical scenes and will identify high frequency action concepts in everyday life, suitable for the implementation of an open set of languages.

1 Introduction

In ordinary language the most frequent action verbs are "general" i.e. they are able to extend to actions belonging to different ontological types (Moneglia & Panunzi 2007). Figure 4 below gives an example of this property. Moreover, each language categorizes action in its own way and therefore the cross-linguistic comparison of verbs denoting everyday activities presents us with a challenging task (Moneglia 2011).

Spontaneous Speech Corpora contain references both to the most frequent actions of everyday life and to their lexical encoding and can be used as a source of semantic information in the domain of an action ontology.

The term Ontology Type is used here to identify the pre-theoretical sets of objects of reference in the domain of Action. Therefore our Ontology will be identified as referring to prototypic eventualities. IMAGACT uses both corpus-based and competence-based methodologies, focusing on high frequency verbs which can provide sufficient variation in spoken corpora. Besides helping in the evaluation of data found in actual language usage, competence based judgments allow us to consider negative evidence which cannot emerge from corpora alone. These judgments are needed to set up cross-linguistic relations. IMAGACT identifies the variation of this lexicon in the BNC-Spoken and, in parallel, in a collection of Italian Spoken corpora (C-ORAL-ROM; LABLITA; LIP; CLIPS). Around 50,000 occurrences of verbs, derived from a 2 million word sampling of both corpora, are annotated.

The project started on March 2011 and involves 15 researchers participating in three main work-packages (Corpus Annotation, Supervision and Cross-linguistic mapping, Validation and Language Extension). The annotation infrastructure is produced by a software house based in Florence (Dr.Wolf srl) and will be delivered as open source.

Roughly 500 verbs per language are taken into account, this represents the basic action oriented verbal lexicon (the Italian part of the task has now been completed, while 50% of the English verbs are still pending). The corpus annotation was performed by three native Italian speaking annotators (with 30 person months devoted to the task) and two native English speaking annotators (13 person months till now).

IMAGACT will result in an Inter-linguistic Action Ontology derived from corpus annotation. Its key innovation is to provide a methodology which exploits the language independent ability to recognize similarities among scenes, distinguishing the *identification* of action types from their *definition*. This ability is exploited both at the corpus annotation level (§2), for mapping verbs of different languages onto the same cross-linguistic ontology (§3) and for validation and extension of the data set to other languages (§4). The paper presents the web infrastructure that has been

developed to this end and the annotation methodology (www.imagact.it/imagact/).

2 Corpus Annotation

The annotation procedure is structured into two main steps: "Standardization & Clustering of Occurrences" and "Types Annotation & Assessment", accomplished by annotators with the assistance of a supervisor. The first task is to examine and interpret verb occurrences in the oral context, which is frequently fragmented and may not provide enough semantic evidence for an immediate interpretation. To this end the infrastructure allows the annotator to read the larger context of the verbal occurrence in order to grasp the meaning (Figure 1 presents one of over 564 occurrences of *to turn* in the corpus). The annotator represents the referred action with a simple sentence in a standard form for easy processing. This sentence must be positively formed, in the third person, present tense, active voice and must fill the essential argument positions of the verb (possible specifiers that are useful in grasping the meaning are placed in square brackets). Basic level expressions (Rosch 1978) are preferred or otherwise a proper name, and word order in sentences must be linear, with no embedding and/or distance relationship.

Crucially, along with the standardization, the annotator assigns the 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 judgement which exploits the annotator's semantic competence (Cresswell 1978) and is given in conjunction with Wittgenstein's hypothesis on how word extensions can be learned (Wittgenstein 1953). The occurrence is judged PRIMARY according to two main operational criteria: 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 turns the idea into a character", as shown in Figure 1 above. We have strong evidence regarding the inter-annotator agreement on this task which may require cross-verification in a few occasions of uncertainty (over 90% in our internal evaluation, based on the performance of two native English and Italian speaking expert annotators).

Only occurrences assigned to the PRIMARY variation class (216 over 564 in this case) make up the set of Action Types stored in the ontology. To this end 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.

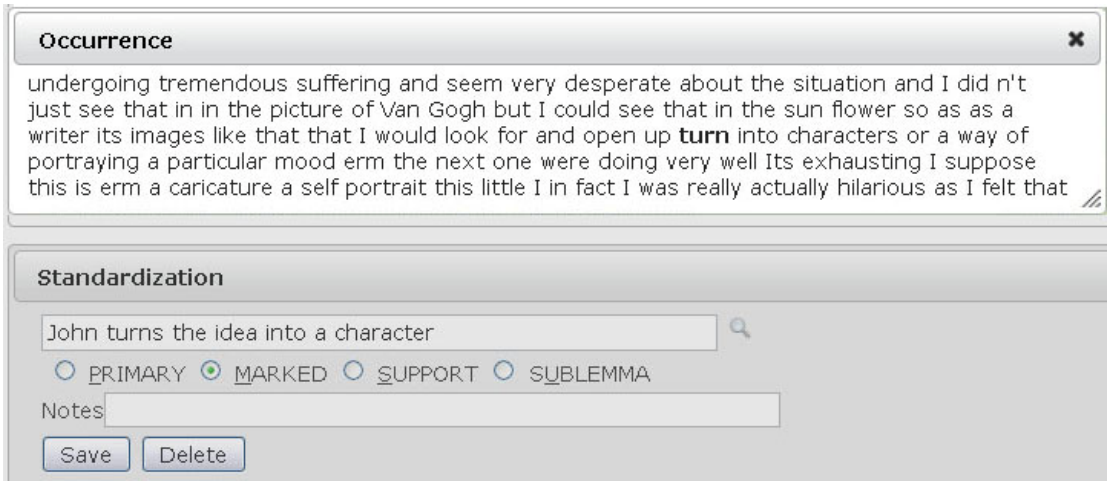

Figure 1. Verb occurrence and Standardization box

The infrastructure is designed to allow the annotator to create types ensuring both cognitive similarity among their events and pragmatic differences between them. The overall criterion for

type creation is to keep granularity to its minimal level, assigning instances 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 (Fellbaum 1998) (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. For instance, standardized occurrences

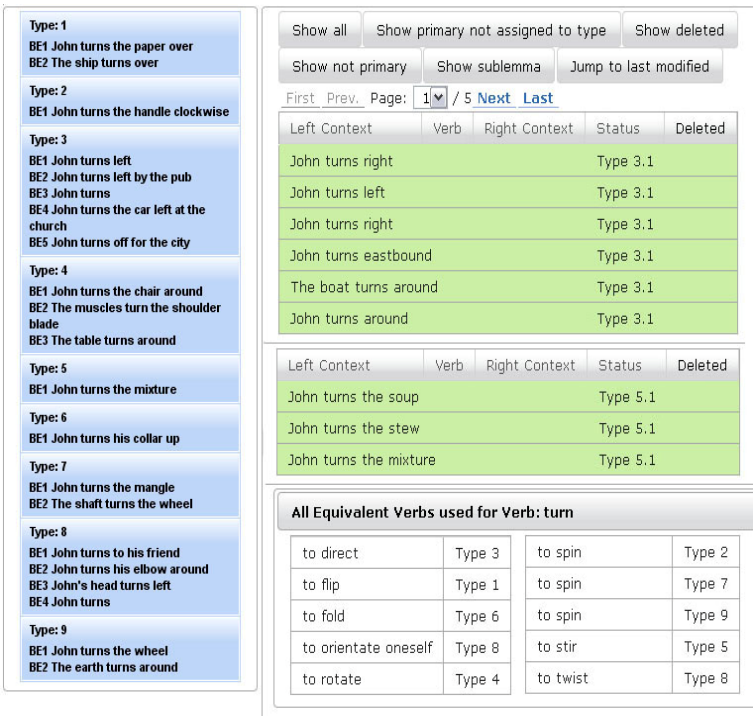

Figure 2 Clustering standardizations into types

of *to turn* are gathered into Type 3 and Type 5 in Figure 2 because all the occurrences can be respectively substituted by *to direct* and *to stir* and the body schema changes from movement into space to an activity on the object.

The infrastructure assists the annotator in the task by showing the types that have been created so far (on the left side) and the equivalent verbs used to differentiate them (at the bottom).

The assigned instances can be shown by type and best example according to the annotator's needs (e.g. Type 3 and Type 5 in the figure). The infrastructure also provides functionality for making easy revisions to hypotheses (show instances not yet assigned, show all instances, verification of Marked variation, editing/merging/splitting types etc.).

The approach underlying the annotation strategy does not require *a priori* any inter-annotator agreement in this core task, which is strongly underdetermined, and rather relies on a supervised process of revision.

Once all occurrences have been processed, the 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 verification criteria are practical: the supervisor verifies that each type cannot be referred to as an instance of another without losing internal cohesion. The operational test checks if it is understandable that the native speaker is referring to the event in *a* by pointing to the prototype in *b*. The supervisor considers the pragmatic relevance of these judgments and keeps the granularity accordingly.

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. Those that aren't are assigned to other types.

The assessment runs in parallel with the annotation of the main linguistic features of a type. More best examples can be added in order to represent all thematic structures of the verb which can satisfy that interpretation. As shown in Figure 3 the thematic grid must be filled, by writing each argument in a separate cell and selecting a role-label from the adjacent combo-box. The tag-set for thematic role annotation is constituted by a restricted set of labels derived from current practices in computational lexicons. We are using Palmer's Tagset in VerbNet[1] with adaptations.

Each best example is also annotated with an aspectual class which is assigned by means of the Imperfective Paradox Test (Dowty, 1979). Aspect can assume three values: event, process or state.

Sentences that are judged peripheral instances of the type can be marked, thus identifying fuzziness in pragmatic boundaries. The annotation procedure ends 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 mapping

Working with data coming from more than one language corpus, IMAGACT must produce a language independent type inventory. For instance, in the case of *to turn* Action types must be consistent with those extended by the Italian verb *girare*, which could be roughly equivalent. Therefore the supervisor will face two lists of types independently derived from corpora annotation. In this scenario, the setting of cross-linguistic relations between verbal entries relies on the identification of a strict similarity between the Types that have been identified (and not through the active writing of a definition). The task is to map similar types onto one prototypical scene that they can be an instance of.

Each prototypical scene is filmed at LABLITA and corresponds to the scripting of one of the best examples selected among all the corpus occurrences which instantiate one Type.

This procedure does not require that the verbs matching onto the same prototypical scene have the same meaning. Two words having different intensions (both within and across languages) may indeed refer to the same action type. The cross-linguistic relation is established accordingly.

Figure 4 roughly sketches the main types derived from the annotation of *to turn* and *girare* and their mapping onto scenes. The supervisor should recognize for instance, that T6 of *girare* and T1 of *to turn* are instances of the same prototype. He will produce one scene accordingly.

The cross-linguistic mapping allows us to predict relevant information which does not emerge from simple corpus annotation. For instance T2 of *girare* never occurs in the English

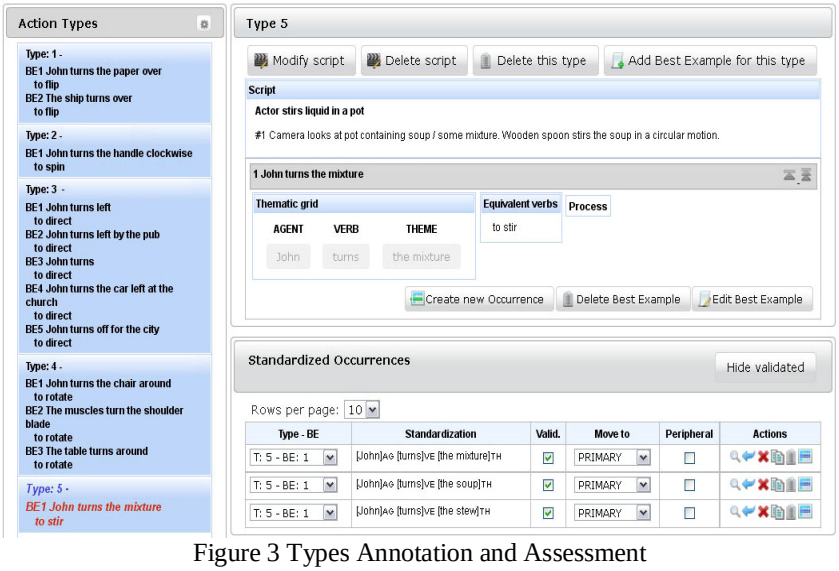

Figure 3 Types Annotation and Assessment

corpus, but native English speakers can recognize from the scene corresponding to T2 that this is also a possible extension of *to turn*. The mapping of the verb onto that type will therefore be established, providing competence based information.

On the contrary, T3 of *girare* and T6 of *to turn* never occur in the English and Italian corpora, however informants recognize that T3 of *girare* cannot be extended by *to turn* (*revolve* is applied) while T6 of *to turn* cannot be extended by *girare* (*alzare* is applied).

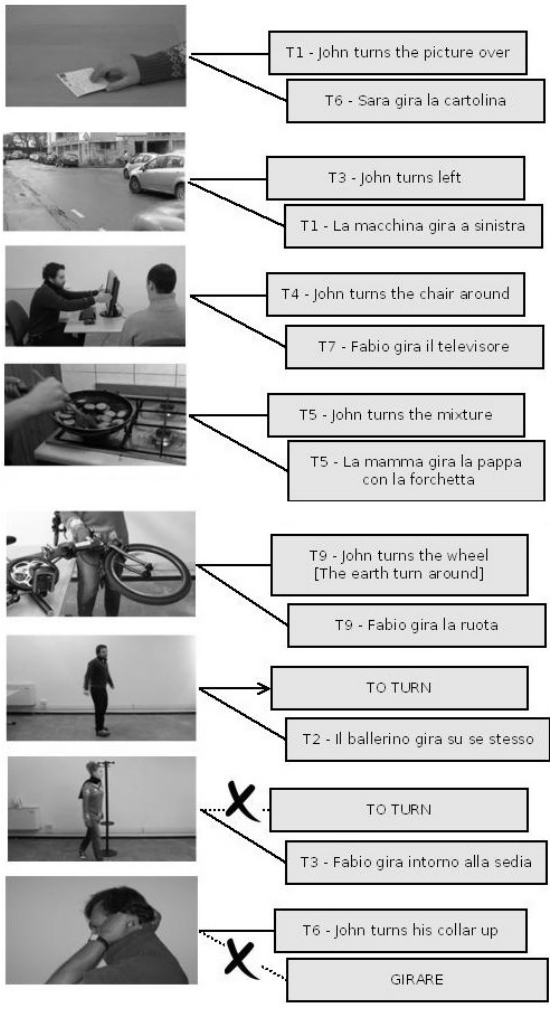

Figure 4. Mapping Action types onto Scenes

In other words the infrastructure and the methodology embodied in it allow the identification of the pragmatic universe of action and of how different languages parse it. This result is obtained in a Wittgenstein-like scenario without the comparison of definitions. The use of prototypical images bypasses this complex problem and permits the identification of the focal pragmatic variation of general verbs and their differentials in different languages.

The link of these scenes to the *sysnets* recorded in WordNet is also carried out when a proper *synset* is available (Moneglia et al. 2012).

Corpora, annotation, lexical variation and cross-linguistic equivalences recorded in each prototypical scene are stored in a database accessed via the web. No annotation format has been so far defined but several current standards in annotation could be relevant here. For the linking between an offset in the corpus and a standardized instance the ISO stand-off annotation format LAF-GrAF could be used. As for the annotation of each standardized instance with syntactic and semantic information (i.e. thematic roles) the ISO MAF and the SemAF could be applicable. Generally speaking, in the framework of the ISO working groups, the IMAGACT annotation procedure as a could be discussed as a possible new work item.

4 Validation and Extension

The direct representation of actions through scenes that can be interpreted independently of language allows the mapping of lexicons from different languages onto the same cross-linguistic ontology. On the basis of this outcome it is possible to ask informants what verb(s) should be applied in his language to each scene and to the set of English and Italian sentences headed by that scene.

Crucially, the informant will verify whether or not the choice is correct for all arguments retrieved from the corpus and assigned to that type and in doing so will verify to which extent the pragmatic concepts stored in the ontology are productive i.e. they permit generalizations at a cross-linguistic level. A concept is valid for cross-linguistic reference to action if, independently of the language, the verb that is applied to the prototypical instance can be also applied to all sentences gathered in it.

The infrastructure organizes this work into two steps: a) alignment of the English and Italian sentences gathered within each entry and generation of a data set of parallel sentences; b) competence based extension (Spanish and Chinese Mandarin). All types in the ontology are checked and all English and Italian action verbs referring to a type will find the appropriate correspondence in

the target languages for that type. The infrastructure allows for the extension to an open set of languages (Moneglia, 2011).

Figure 5 is an example of a competence based extension to Chinese for what regards the second and first scenes of Figure 4. The infrastructure: a) presents the set of sentences gathered into one scene; b) requests the user to input a verb in the target language; c) asks whether or not this verb can be applied in all gathered sentences. The Chinese informant verified that the two scenes require two different verbs (*zhuǎn* and *fān*) which were appropriate in all occurrences.

Distinguishing families of usages of general verbs from the granular variations allows us to establish productive cross-linguistic relations, so validating the Ontology entries in the real world.

Figure 5 Validation & Extension interface

References

British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services URL: http://www.natcorp.ox.ac.uk/

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

C-ORALROM http://catalog.elra.info/product_info.php?products_id=757

Cresswell M. F. 1978 Semantic Competence in F. Guenthner, M. Guenthner-Reutter, Meaning and translation. NY University Press: New York, 9-28

De Mauro T., Mancini F., Vedovelli M., Voghera M. 1993. Lessico di frequenza dell'italiano parlato (LIP). Milano: ETASLIBRI.

Dowty, D. 1979. Word meaning and Montague grammar. Dordrecht: Reidel.

Fellbaum, Ch. (ed.) 1998. WordNet: An Electronic Lexical Database. Cambridge: MIT Press.

Ide, N. and K, Suderman. 2007.. "GrAF: A graph-based format for linguistic annotations". In *Proceedings of the Linguistic Annotation Workshop at ACL 2007*. Prague, Czech Republic**:** 1-8.

International Organization for Standardization. 2012. ISO DIS 24612- Language Resource Management - Linguistic annotation framework (LAF). ISO/TC 37/SC4/WG 2.

International Organization for Standardization. 2008. ISO DIS 24611 Language Resource Management - Morpho-syntactic Annotation Framework (MAF). ISO/TC 37/SC4/WG 2.

International Organization for Standardization. 2008. ISO DIS 24617- Language Resource Management - Semantic annotation framework (SemAF). ISO/TC 37/SC4/WG 2.

Levin, B. 1993. English verb classes and alternations: A preliminary investigation. Chicago: University of Chicago Press.

Moneglia M. 2011. Natural Language Ontology of Action. A gap with huge consequences for Natural Language Understanding and Machine Translation, in in Z. Vetulani (ed.)Human Language Technologies as a Challenge for Computer Science and Linguistics. Poznań: Fundacja Uniwersytetu im. A. Mickiewicza 95-100.

Moneglia, M., Monachini, M., Panunzi, A., Frontini, F., Gagliardi, G., Russo I. 2012 Mapping a corpus-induced ontology of action verbs on ItalWordNet. In C. Fellbaum, P. Vossen (eds) Proceedings of the 6th International Global WordNet Conference (GWC2012) Brno. 219-226.

Moneglia, M. & Panunzi, A., 2007. Action Predicates and the Ontology of Action across Spoken Corpora. In: M. Alcántara & T. Declerck, Proceeding of SRSL7. Universidad de Salamanca, 51-58.

Rosch, E. 1978. Principles of Categorization. In E. Rosch & B.B. Lloyd (eds), Cognition and Categorization. Hillsdale: Lawrence Erlbaum, 27–48.

VerbNet http://verbs.colorado.edu/~mpalmer/projects/verbnet.html

Wittgenstein, L. 1953. Philosophical Investigations. Oxford: Blackwell.