10.12871/CLICIT2014126

Evaluating ImagAct-WordNet mapping for English and Italian through videos

Irene De Felice, Roberto Bartolini, Irene Russo, Valeria Quochi, Monica Monachini
Istituto di Linguistica Computazione A. Zampolli, ILC CNR Pisa
firstname.lastname@ilc.cnr.it

Abstract

English. In this paper we present the results of the evaluation of an automatic mapping between two lexical resources, WordNet/ItalWordNet and ImagAct, a conceptual ontology of action types instantiated by video scenes. Results are compared with those obtained from a previous experiment performed only on Italian data. Differences between the two evaluation strategies, as well as between the quality of the mappings for the two languages considered in this paper, are discussed.

Italiano. *L'articolo presenta i risultati della valutazione di un mapping automatico realizzato tra due risorse lessicali, WordNet/ItalWordNet e ImagAct, un'ontologia concettuale di tipi azionali rappresentati per mezzo di video. Tali risultati vengono confrontati con quelli ottenuti da un precedente esperimento, condotto esclusivamente sull'italiano. Vengono inoltre discusse le differenze tra le due strategie di valutazione, così come nella qualità del mapping proposto per le due lingue qui considerate.*

1 Introduction

In lexicography, the meaning of words is represented through words: definitions in dictionaries try to make clear the denotation of lemmas, reporting examples of linguistic usages that are fundamental especially for function words like prepositions. Corpus linguistics derives definitions from a huge amount of data. This operation improves words meaning induction and refinements, but still supports the view that words can be defined by words.

In the last 20 years dictionaries and lexicographic resources such as WordNet have been enriched with multimodal content (e.g. illustrations, pictures, animations, videos, audio files). Visual representations of denotative words like concrete nouns are effective: see for example the ImageNet project, that enriches WordNets glosses with pictures taken from the web.

Conveying the meaning of action verbs with static representations is not possible; for such cases the use of animations and videos has been proposed (Lew 2010). Short videos depicting basic actions can support the users need (especially in second language acquisition) to understand the range of applicability of verbs. In this paper we describe the multimodal enrichment of ItalWordNet and WordNet 3.0 action verbs entries by means of an automatic mapping with ImagAct (www.imagact.it), a conceptual ontology of action types instantiated by video scenes (Moneglia et al. 2012). Through the connection between synsets and videos we want to illustrate the meaning described by glosses, specifying when the video represents a more specific or a more generic action with respect to the one described by the gloss. We evaluate the mapping watching videos and then finding out which, among the synsets related to the video, is the best to describe the action performed.

2 ImagAct and ItalWordNet/WordNet: general principles

In ImagAct, concrete verbs meanings are represented as 3D videos and, from a theoretical point of view, different meanings of the same verb are intended as different conceptual basic action types. ImagAct action types have been derived bottom-up, by annotating occurrences of 600 high frequency Italian and English action verbs, previously extracted from spoken corpora. All occurrences have been manually clustered into action types, on the basis of body movements and objects involved. Each lemma usually has more than one action type: for example, for the verb to open we have 7 basic action types, each of them denoting a different physical action and applicable to different sets of objects. This process was carried out in parallel on English and Italian data; finally, Italian and English action types were mapped onto one another and refinements or adjustments were made in order to stabilise the ontology. In this way, 1100 basic action types have been identified.

The ontologys nodes (action types) consist of videos created as 3D animations, each one provided with the sentence that best exemplifies it, according to annotators; each short video represents a particular type of action (e.g. a man taking a glass from a table) and it is related to a list of Italian and English verbs that can be used to describe that action (all the lemmas associated to a scene can thus be seen as something quite similar to WordNet synsets). The 3D animations represent the gist of an action in terms of movements and interactions with the object in a pragmatically neutral context. Sometimes, high level actional concepts could not be represented with a video: in this case, an ontological node is created and associated to a scene ID as well as to a list of Italian and English verbs, but no video is uploaded in the resource. This said, it is evident that ImagAct is a lexical resource structured in a multimodal way: videos represent the core of the resource.

If in WordNet (Fellbaum 1998) and in ItalWordNet (Roventini et al. 2000) lexicographic principles guide the individuation of meanings, ImagAct aims to list the different concepts (one or more) which we refer to when using action verbs. Furthermore, WordNet aims to describe all different uses of a verb, including idiomatic or metaphorical expressions, whereas ImagAct is specifically focused on linguistic uses related to concrete actions. Being aware of the differences between the two resources, we want to map ImagAct on WordNet not only to make clear how the focus on the perceptual aspect of actions can cause the induction of different verbs' senses, but also to enrich WordNet with videos depicting the actions denoted by glosses.

3 Methods

We describe an approach inspired by ontology matching methods for the automatic mapping of ImagAct video scenes onto WordNet/ItalWordNet. The aim of the mapping is to automatically establish correspondences between WN verbal synsets and ImagAct basic action types. This can be done by measuring the semantic proximity between video scenes and synsets in terms of overlap between the class of verbs (lemmas) associated to a scene in ImagAct and the set of synonyms in WordNet synsets (together with their hypernyms and hyponyms).

The ImagAct dataset used for the mapping consists of 1120 video scenes, with a total of 1100 associated Italian verb types (500 lemmas, with an average of 2.4 verb lemmas per scene). For English, we have 1163 video scenes, with a total of 1181 associated English verb types (543 lemmas, with an average of 2.2 verb lemmas per scene). The difference between Italian and English number of scenes is due to the fact that some action types have only been identified for English and cannot be mapped on any Italian action types.

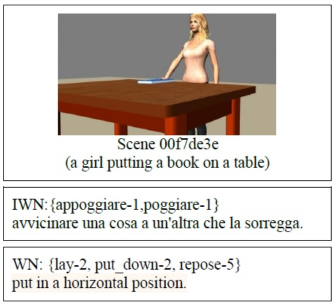
Concerning WordNet, we consider as relevant information: verbal synsets, verb senses, hyponymic and hypernymic relations. Altogether, the ItalWordnet database (hosted at CNR-ILC) contains 8903 verbal synsets and 14086 verb senses (8121 lemmas, with an average of 1.1 verb lemmas per synset) that are potential candidates for the mapping.

As described in Bartolini et al. (2014), we implemented an algorithm inspired by Rodriguez and Egenhofer (2003), based on set-theory and feature-based similarity assessment (Jaccard, 1912; Tversky, 1977), which proved particularly interesting for the mapping of different and independent ontologies and especially fit for lexical resources, as it is primarily based on word matching (for details about the mapping algorithm, see Bartolini et al., 2014). In that paper we presented the mapping between ImagAct and ItalWordNet. The evaluation was performed on a gold standard of 260 Italian verb lemmas corresponding to 358 action types, which mapped onto a total of 343 ItalWordNet synsets. This gold standard was created by mapping verb action types (not scenes) to ItalWordNet synsets. The performance of the algorithm was assessed on the same task of mapping verb types onto synsets: a similarity score was calculated between the verbs contained in a synset and those related to an action type; the best candidate synset is thus the synset with the bigger overlap with the action type, as this overlapping is measured by the algorithm. In terms of performance, our evaluation results (recall 0.61, precision 0.69) proved that, at least for WordNet-like lexical resources, differences in the synonym sets are relevant for assessing the proximity or distance of concepts.

Since results from this first experiment were encouraging, we adopted the same algorithm also for the English mapping (ImagAct-Princeton WordNet). The database of WordNet 3.0 contains 13767 verbal synsets and 25047 verb senses (11529 lemmas, with an average of 1,19 verb lemmas per synset), as potential candidates for the mapping. For this task we had no gold standard previously created, thus a new evaluation strategy was assessed and then conducted on both English and Italian data. We think that this will not only improve the judgment of the quality of the mapping proposed, but also allow us to compare results from two different kinds of evaluating methods.

4 Evaluation

In this paper we propose a new evaluation of both the English and the Italian mapping. The evaluation was conducted by two authors, respectively on Italian and English data. To test the quality of the mapping proposed by the algorithm, we decided to select a group of ImagAct scenes related to the actions of putting and then to manually assign a judgement to the definitions of the candidate synsets proposed for the mapping for both languages. The two steps of the evaluation were carried out in parallel, one that considers the mapping proposed between ImagAct scenes and ItalWordNet synsets, and the other that considers the mapping proposed between the same ImagAct scenes and English WordNet synsets.



Scene 00f7de3e
(a girl putting a book on a table)

IWN: {appoggiare-1,poggiare-1}
avvicinare una cosa a un'altra che la sorregga.

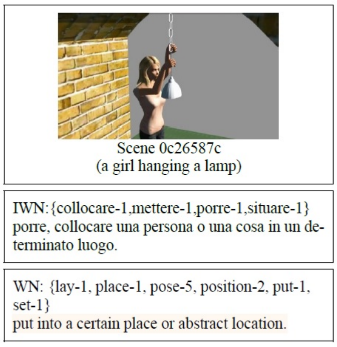
WN: {lay-2, put_down-2, repose-5}
put in a horizontal position.

Figure 1: Examples for Imagact-(Ital)WordNet mapping evaluation: equivalence relation.

We expected four possible cases of acceptable mapping (for each one we report, when possible, examples from the two languages):

1. The synset's gloss perfectly describes the scene (equivalence relation (see Figure 1).

2. The synsets gloss describes an event that is more general than that represented by the scene (WordNet more generic).



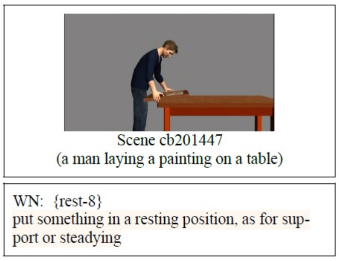
Scene 0c26587c
(a girl hanging a lamp)

IWN: {collocare-1,mettere-1,porre-1,situare-1}
porre, collocare una persona o una cosa in un determinato luogo.

WN: {lay-1, place-1, pose-5, position-2, put-1, set-1}
put into a certain place or abstract location.

Figure 2: Examples for ImagAct-(Ital)WordNet mapping evaluation: WordNet more generic.

3. The synsets gloss describes an event that is more specific than that represented by the scene (WordNet more specific).



Scene cb201447
(a man laying a painting on a table)

WN: {rest-8}
put something in a resting position, as for support or steadying

Figure 3: Examples for Imagact-(Ital)WordNet mapping evaluation: WordNet more specific.

4. The synset's gloss is unrelated to the scene (no relation).

Details about the evaluation are reported in Table 1.

	Scene (tot.)	Scene without videos
IT	108	27
EN	111	29

Table 1: Scenes evaluated.

The difference in terms of scenes between Italian and English depends on the fact that it is possible that one scene is pointed to only by English verbs, thus this scene cannot be mapped on ItalWordNet.

The results of the evaluation are summarised in the Table 2: in each column is reported the number of scenes that can be described exactly (=) with the gloss of the first, second or third synset in the mapping, Considering that for each scene the average number of mapped synsets is 60 for Italian and 65 for English, and that we chose to evaluate a group of scenes representing actions that include very generic verbs such as to put and to bring, results for Italian are very good: in the vast majority of cases the right synset is among the first three synsets evaluated as appropriate by the mapping algorithm. Only in the 14.8% of cases no possible match was found. The main factor that impacts on the results for English depends on the way WordNet is structured: in WordNet we find more synonyms with respect to ItalWordNet and as a consequence the mapping algorithm has a different performance. An example of the mapping resulted from the evaluation is available at http://tinyurl.com/q32cps6.

	Italian		English	
	=	all	=	all
First result	41	64	15	33
Second result	2	4	2	5
Third result	1	1	2	5
All	69 (85.2%)		43 (52.4%)	

Table 2: Evaluation results.

5 Conclusions

Mutual enrichments of lexical resources is convenient, especially when different kinds of information are available. In this paper we describe the mapping between ImagAct videos representing action verbs' meanings and WordNet/ItalWordNet, in order to enrich the glosses multimodally. Two types of evaluation have been performed, one based on a gold standard that establishes correspondences between ImagActs basic action types and ItalWordNets synsets (Bartolini at al. 2014) and the other one based on the suitability of a synsets gloss to describe the action watched in the videos. The second type of evaluation suggests that for Italian the automatic mapping is effective in projecting the videos on ItalWordNet's glosses. For what regards the mapping for English, as future work we plan to change the settings, in order to test if the number of synonyms available in WordNet has a negative impact on the quality of the mapping.

Acknowledgments

This research was supported by the MODELACT project, funded by the Futuro in Ricerca 2012 programme (Project Code RBFR12C608); website: http://modelact.lablita.it.

References

Roberto Bartolini, Valeria Quochi, Irene De Felice, Irene Russo, and Monica Monachini. 2014. From Synsets to Videos: Enriching ItalWordNet Multimodally. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, 3110-3117.

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

Jaccard Paul, 1912. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37-50.

Massimo Moneglia, Gloria Gagliardi, Alessandro Panunzi, Francesca Frontini, Irene Russo, and Monica Monachini. 2012. IMAGACT: Deriving an Action Ontology from Spoken Corpora. In *Proceedings of the Eighth Joint ACL - ISO Work- shop on Interoperable Semantic Annotation*, 42-47.

Andrea M. Rodrguez and Max J. Egenhofer. 2003. Determining Semantic Similarity among Entity Classes from Different Ontologies. In *IEEE Transactions on Knowledge and Data Engineering*, 15(2): 442-456..

Adriana Roventini, Antonietta Alonge, Nicoletta Calzolari, Bernardo Magnini, and Francesca Bertagna. 2000. ItalWordNet: a Large Semantic Database for Italian. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, 783-790.

Amos Tversky. 1977. Features of similarity. *Psychological Review*, 84(4):327-352.