Enriching BabelNet verbal entities with videos: a linking experiment with the IMAGACT ontology of action

Lorenzo Gregori
University of Florence
lorenzo.gregori@unifi.it

Andrea Amelio Ravelli
University of Florence
aramelior@gmail.com

Alessandro Panunzi
University of Florence
alessandro.panunzi@unifi.it

Abstract

Herein we present a study dealing with the linking of the two multilingual and multimedia resources BabelNet and IMAGACT, which seeks to connect the videos contained in the IMAGACT Ontology of Actions with related verb entries in BabelNet. The linking experiment is based on an algorithm that exploits the lexical information of the two resources and preliminary results show that is possible to achieve extensive linking between the two. This linkage is highly desirable for both of the resources as the BabelNet verbal entries may be enriched with the visual representations found in IMAGACT and the IMAGACT ontology can be connected to the BabelNet semantic network. A connection between the two resources yields a knowledge base that may be exploited for complex tasks such as reference disambiguation and the automatic or assisted translation of both verbs and sentences which refer to actions.

1 Introduction

Ontologies are widely used to represent language resources on the web and allow them to be exploited within natural language processing tasks. The availability of formal representation languages like RDF and the growing development of top-level ontologies like *lemon* (McCrae et al., 2011) are leading toward a unified method for publishing lexico-semantic resources as open data. Following in this direction, the Linguistic Linked Open Data Cloud (Chiarcos et al., 2011) presently collects more than 500 language resources in RDF.

Data interconnection is a fundamental aspect of natural language processing, as evidenced by the increasing research into mapping and linking techniques among ontologies (Otero-Cerdeira et al., 2015). Among the challenges to be overcome are differences between the resources, which are usually built using their own individual criteria, and growing language datasets, which necessitate automatic procedures since manual approaches are costly and unscalable (Siemoneit et al., 2015). Instance matching techniques play an important role in this context, because they provide a way to link resources without having to map ontological entities (Castano et al., 2008; Nath et al., 2014).

This paper presents a linking hypothesis between BabelNet (Navigli and Ponzetto, 2012a) and IMAGACT (Moneglia et al., 2014a), two multilanguage and multimedia resources both suitable for automatic translation and disambiguation tasks (Moro and Navigli, 2015; Russo et al., 2013; Moneglia, 2014). In contrast with previous attempts to map IMAGACT and WordNet (De Felice et al., 2014; Bartolini et al., 2014), we propose switching to a multilingual viewpoint and implementing light linking between the resources as opposed to a full-scale mapping. This seems to be a more fruitful approach, especially considering the differences between a wide-ranging knowledge base such as BabelNet and a specific action ontology like IMAGACT, which aren't very compatibile in terms of the definition concepts (Gregori, 2016).

2 Resources

2.1 BabelNet

BabelNet[1] is a multilingual semantic network created from the mapping together of two of the most important web resources available: the WordNet thesaurus and the Wikipedia enciclopedia. At present, BabelNet 3.5 contains 272 languages and is among the widest multilingual resources available for semantic disambiguation. Semantic and encyclopedic data have been grouped together and linked by way of an automatic mapping algorithm

[1] http://babelnet.org

so as to create a dictionary with a wealth of information and a dense network of semantic relations. Concepts and entities are represented by BabelSynsets (BS), extensions of WordNet synsets: a BS is a unitary concept identified by semantic features, glosses and usage examples and lemmas in different languages refer to it. BabelNet has grown much in the past 3 years and received large contributions from its mapping together with resources such as ImageNet, GeoNames, and OmegaWiki (along with many others), which increased its information beyond just the lexical data and produced a wide-ranging, multimedia knowledge base.

2.2 IMAGACT

IMAGACT[2] is a visual ontology of action developed within the Italian PAR/FAS 2007-2013 Program that provides an image-based translation and disambiguation framework for general verbs in 4 languages: Italian, English, Chinese, and Spanish. The database has been evolving under the MODELACT Project[3] (Futuro in Ricerca 2012) since its first release and at present contains 9 fully-mapped languages and 13 which are underway (Moneglia et al., 2014b). The resource is built on an ontology containing a fine-grained categorization of action concepts, each represented by one or more action prototypes in 3D animation format. IMAGACT currently contains 1,010 scenes which encompass the action concepts most commonly referred to in everyday language usage. The data is derived from the detailed annotation of verb occurrences taken from spoken corpora (Moneglia et al., 2012).

The links between the verbs and video scenes are based on local equivalences between different verbs within a context expressed by a scene (i.e. the ability of the different verbs to describe the same action, visualised in the scene). Each concept is represented by one or more scenes: specific concepts categorize pragmatically similar actions and can be identified by a single prototype, while general concepts cover a wider variety of actions and require a family of prototypes to be clearly expressed. These visual representations convey the action information in a cross-linguistic environment and the IMAGACT ontology may thus be exploited for reference disambiguation, being of particular use in automatic and assisted translation tasks (Panunzi et al., 2014).

3 Linking experiment

The aim of this experiment is to link the IMAGACT video scenes to the BabelNet interlinguistic concepts (BabelSynsets). In fact, the BabelNet objects are already enriched with visual information, though this information contains static images only which are usually inadequate for representing action concepts. In this way, adding video scenes to the verbs is very desirable and would suggest itself as a natural extension of BabelNet. IMAGACT, similarly, stands to increase and improve its lexical information by acquiring the lemmas connected to the Babel synsets.

The linking process is based on an algorithm that exploits the lexical information of the two resources. It performs a comparison between the BabelNet semantic network and the IMAGACT ontology so as to automatically identify the IMAGACT action scenes that may serve as adequate prototypes for the corresponding Babel synsets[4].

Table 1 reports on the 17 languages common to both BabelNet and IMAGACT, detailing the relative number of verbs in each, and constitutes the quantitative data on which a matching algorithm may be run.

3.1 The test set

A manually annotated dataset of 25 scenes and 30 Babel synsets (750 judgments) was created in order to test the algorithm and evaluate the results. Both the scenes and the Babel synsets belong to the semantic variation of 7 English verbs, all of which have marked ambiguity and high frequency in the language usage: *put*, *move*, *take*, *insert*, *press*, *give* and *strike*.

To obtain a representative sampling we needed to consider some relevant aspects regarding the number of relations between the verbs and the scenes. First of all, since we are dealing with a still developing ontology the languages in IMAGACT are not all represented equally, and the same is true of BabelNet: essentially, we have a different number of verbs belonging to each language (see Table 1). Then, the number of verbs that can be used to refer to an action within a language is highly dependent on the action itself: some scenes can be described by several verbs and are highly linked,

[2] http://www.imagact.it

[3] http://modelact.lablita.it/

[4] This test is based on BabelNet 3.0; the data was extracted using the Java API (Navigli and Ponzetto, 2012b)

Language	BN Verbs	IM Verbs
English (EN)	29,738	1,299
Polish (PL)	9,660	1,145
Chinese (ZH)	9,507	1,171
Italian (IT)	6,686	1,100
Spanish (ES)	6,159	736
Russian (RU)	4,975	36
Portuguese (PT)	4,624	776
Arabian (AR)	4,216	331
German (DE)	3,754	992
Norwegian (NO)	1,729	115
Danish (DA)	1,685	588
Hebrew (HE)	1,647	162
Serbian (SR)	858	1,178
Hindi (HI)	831	220
Urdu (UR)	233	83
Sanskrit (SA)	33	225
Oriya (OR)	6	178
Total	86,435	10,481

Table 1: The 17 shared languages of BabelNet (BN) and IMAGACT (IM) with verbal lemma counts.

while other scenes offer much lower predication possibilities. Lastly, since each language categorizes the action space in its own way, the same action may be referred to by many verbs in one language and by just a few in another.

By taking these aspects into account we used a sampling strategy that preserved the numeric variability of the verb-to-scene relations; the samples were randomly selected with this constraint in mind. We had a similar quantitative discrepancy for the verb-to-BS relations and we used a similar sampling strategy to select the 30 BS candidates.

Each ⟨BS,Scene⟩ pair has been evaluated to check if the scene is appropriate in representing the BS. Three annotators compiled the binary judgment table and we reported the values shared by at least 2 of 3. The measured Fleiss' kappa inter-rater agreement for this task was 0.76 [5].

A bigger test set is currently being created in order to fine tune the parameters and increase the algorithm's accuracy, at which point it will be run on all 1,010 IMAGACT scenes.

[5]The annotated test set is published at `http://bit.ly/1MtZqB9`

3.2 Algorithms

The basic algorithm that we used for this experiment is based on a function that compares a scene with a BS and calculates the frequency whereby the verbs connected to the scene belong the the BS too. The set of candidates are all the possible BS for each verb connected to the scene.

Algorithm 1 Basic algorithm

1: s: input scene
2: V: set of verbs linked to s in IMAGACT
3: List⟨BabelSynset⟩ LS: empty list
4: for each v_i in V
5: List⟨BabelSynset⟩ Syn = list of BS connected to v_i
6: add Syn to LS
7: List⟨BabelSynset⟩ $FLS = freqList(LS)$
8: Connect s to the first n BabelSynset in FLS

The *freqList* function computes the frequency list of the BSs in *LS* and sorts them according to the most frequent.

Starting from this basic linking algorithm, we implemented an improved version of it, in which the BabelNet semantic network is exploited: this allowed us to also consider the BSs that are semantically related to the target one. We defined a recursive function w that weights the impacts of the BSs that are connected to the main one.

Given S the set of BSs, $s_0 \in S$ directly connected to the verb, and $s', s'' \in S$ connected through a relation $r \in R$, we defined a function $w : S \rightarrow [0, 1]$ so that

- $w(s_0) = freq(s_0)$
- $w(s'') = w(s') \cdot c \cdot p(r)$

where

- $freq$ computes the BS frequency as in the basic algorithm;
- R is the set of relations between the verbal BSs;
- $p(rel) : R \rightarrow [0, 1]$ is a function that assigns a different weight to each relation R;
- $c \in [0, 1]$ is a coefficient that decreases the weight when the distance from the main node increases.

Using this metric we are able to tune the impact of the different BabelNet semantic relations. In fact, we proved that some of them are very relevant for this task while others convey non-pertinent information and should be excluded. Table 2 shows the list of relations between the verbal BSs and their relevance values, measured with Information Gain on the annotated dataset.

BabelNet relations	***IG* value**
Hyponym	0.135
Also See	0.050
Hypernym	0.041
Verb Group	0.039
Entailment	0.009
Gloss Related	0.000
Antonym	0.000
Cause	0.000

Table 2: Relations between verbal BSs.

3.3 Results

The 2 algorithms were run on the test set and the accuracy of the first n in selecting the BSs for the scenes was verified. Table 3 gives a result summary, while the complete result set is available at `http://bit.ly/1MtZqB9`.

	Basic Alg.	**Improved Alg.**
% corr. $(n = 1)$	100%	100%
% corr. $(n = 2)$	84%	88%
% corr. $(n = 3)$	76%	83%

Table 3: Percentage of correct assignments of BSs to scenes for the 2 algorithms with different n values.

In both algorithms the first selected BS is always correct. The results worsen by increasing the value of n while at the same time the qualitative gap between the 2 algorithms increases: this proves that the information conveyed by the neighboring nodes in the semantic graph is relevant to this task.

4 What's next

The results obtained with this first linking experiment between IMAGACT and BabelNet are promising and we are building a bigger test set in order to fine-tune the parameters. Running the algorithm on the whole set of scenes is an important step; in particular we need to better estimate the impact of each BabelNet relation ($p(rel)$ function) and the impact of the neighboring nodes (c coefficient).

At present the number of BSs that one scene is going to be linked to (the n parameter of the algorithm) is fixed for every scene. Considering the high numeric difference in verb reference possibilities in an inter-linguistic context (see 3.1), this approach is not appropriate. Instead we need to determine a threshold for the weighting function that can be used to decide if the similarity score of each pair ⟨BS,Scene⟩ is enough to establish a link.

Finally, we feel it important to note that this procedure is scalable and the algorithm's accuracy should naturally increase with growth in IMAGACT's languages and lemmas, since the matching strategy is based on overlap between lexical entries.

References

Roberto Bartolini, Valeria Quochi, Irene De Felice, Irene Russo, and Monica Monachini. 2014. From synsets to videos: Enriching italwordnet multimodally. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

S. Castano, A. Ferrara, D. Lorusso, and S. Montanelli. 2008. On the ontology instance matching problem. In *Database and Expert Systems Application, 2008. DEXA '08. 19th International Workshop on*, pages 180–184, Sept.

Christian Chiarcos, Sebastian Hellmann, and Sebastian Nordhoff. 2011. Towards a linguistic linked open data cloud: The open linguistics working group. *TAL*, pages 245–275.

Irene De Felice, Roberto Bartolini, Irene Russo, Valeria Quochi, and Monica Monachini. 2014. Evaluating ImagAct-WordNet mapping for English and Italian through videos. In Roberto Basili, Alessandro Lenci, and Bernardo Magnini, editors, *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & the Fourth International Workshop EVALITA 2014*, volume I, pages 128–131. Pisa University Press.

Lorenzo Gregori. 2016. *La forma dellontologia dellazione IMAGACT: dal modello gerarchico al modello insiemistico in un DB a grafo*. Ph.D. thesis, University of Florence.

John McCrae, Dennis Spohr, and Philipp Cimiano. 2011. Linking Lexical Resources and Ontologies on the Semantic Web with lemon. In *Proceedings of the 8th Extended Semantic Web Conference on The Semantic Web: Research and Applications - Volume Part I*, ESWC'11, pages 245–259, Berlin, Heidelberg. Springer-Verlag.

Massimo Moneglia, Francesca Frontini, Gloria Gagliardi, Irene Russo, Alessandro Panunzi, and Monica Monachini. 2012. Imagact: deriving an action ontology from spoken corpora. *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (isa-8)*, pages 42–47.

Massimo Moneglia, Susan Brown, Francesca Frontini, Gloria Gagliardi, Fahad Khan, Monica Monachini, and Alessandro Panunzi. 2014a. The IMAGACT Visual Ontology. An Extendable Multilingual Infrastructure for the Representation of Lexical Encoding of Action. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Massimo Moneglia, Susan Brown, Aniruddha Kar, Anand Kumar, Atul Kumar Ojha, Heliana Mello, Niharika, Girish Nath Jha, Bhaskar Ray, and Annu Sharma. 2014b. Mapping Indian Languages onto the IMAGACT Visual Ontology of Action. In Girish Nath Jha, Kalika Bali, Sobha L, and Esha Banerjee, editors, *Proceedings of WILDRE2 - 2nd Workshop on Indian Language Data: Resources and Evaluation at LREC'14*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Massimo Moneglia. 2014. Natural Language Ontology of Action: A Gap with Huge Consequences for Natural Language Understanding and Machine Translation. In Zygmunt Vetulani and Joseph Mariani, editors, *Human Language Technology Challenges for Computer Science and Linguistics*, volume 8387 of *Lecture Notes in Computer Science*, pages 379–395. Springer International Publishing.

Andrea Moro and Roberto Navigli. 2015. SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 288–297, Denver, Colorado, June. Association for Computational Linguistics.

Rudra Nath, Hanif Seddiqui, and Masaki Aono. 2014. An efficient and scalable approach for ontology instance matching. *Journal of Computers*, 9(8).

Roberto Navigli and Simone Paolo Ponzetto. 2012a. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Roberto Navigli and Simone Paolo Ponzetto. 2012b. Multilingual WSD with just a few lines of code: the BabelNet API. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, Jeju, Korea.

Lorena Otero-Cerdeira, Francisco J. Rodrguez-Martnez, and Alma Gmez-Rodrguez. 2015. Ontology matching: A literature review. *Expert Systems with Applications*, 42(2):949 – 971.

Alessandro Panunzi, Irene De Felice, Lorenzo Gregori, Stefano Jacoviello, Monica Monachini, Massimo Moneglia, Valeria Quochi, and Irene Russo. 2014. Translating Action Verbs using a Dictionary of Images: the IMAGACT Ontology. In *XVI EURALEX International Congress: The User in Focus*, pages 1163–1170, Bolzano / Bozen, 7/2014. EURALEX 2014, EURALEX 2014.

Irene Russo, Francesca Frontini, Irene De Felice, Fahad Khan, and Monica Monachini. 2013. Disambiguation of Basic Action Types through Nouns Telic Qualia. In Roser Saur, Nicoletta Calzolari, Chu-Ren Huang, Alessandro Lenci, Monica Monachini, and James Pustejovsky, editors, *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon. Generative Lexicon and Distributional Semantics*, pages 70–75.

Benjamin Siemoneit, John Philip McCrae, and Philipp Cimiano. 2015. Linking four heterogeneous language resources as linked data. In *Proceedings of the 4th Workshop on Linked Data in Linguistics: Resources and Applications*, pages 59–63, Beijing, China, July. Association for Computational Linguistics.