

# Linking dei contenuti multimediali tra ontologie multilingui: i verbi di azione tra IMAGACT e BabelNet

Lorenzo Gregori, Andrea Amelio Ravelli, Alessandro Panunzi

Università di Firenze

{lorenzo.gregori, alessandro.panunzi}@unifi.it,  
aramelior@gmail.com

## Abstract

**English.** We present a study dealing with the linking between two multilingual and multimedial resources, BabelNet and IMAGACT. The task aims to connect the videos contained in the IMAGACT Ontology of Actions and the related verb entries in BabelNet. The linking experiment is based on an algorithm that exploits the lexical information of the two resources. The results show that is possible to achieve an extensive linking between the two ontologies. This linking is highly desirable in order to build a rich multimedial knowledge base that can be exploited for the following complex tasks: the reference disambiguation and the automatic/assisted translation of both the verbs and the sentences which refer to actions.

**Italiano.** Lo studio qui presentato riguarda il collegamento tra due risorse multilingui e multimediali, BabelNet e IMAGACT. In particolare, l'esperimento di linking ha come oggetto i video dell'ontologia dell'azione IMAGACT e le rispettive entrate lessicali verbali di BabelNet. Il task è stato eseguito attraverso un algoritmo che opera sulla base delle informazioni lessicali presenti nelle due risorse. I risultati del linking mostrano che è possibile effettuare un collegamento estensivo tra le due ontologie. Tale collegamento è auspicabile nel senso di fornire una base di dati ricca e multimediale per i complessi task di disambiguazione del riferimento dei verbi di azione e di traduzione automatica e assistita delle frasi che li contengono.

## 1 Introduzione<sup>1</sup>

Le ontologie sono strumenti ampiamente utilizzati per rappresentare risorse linguistiche sul web e renderle sfruttabili da metodi di elaborazione automatica del linguaggio naturale. La disponibilità di linguaggi formali condivisi, come RDF e OWL, e lo sviluppo di ontologie di alto livello, come *lemon* (McCrae et al., 2011), stanno portando ad una metodologia unificata per la pubblicazione delle risorse linguistiche in forma di *open data*<sup>2</sup>.

La rappresentazione delle informazioni attraverso le ontologie non è però sufficiente alla costruzione della rete semantica che è alla base dei nuovi paradigmi del web. L'interconnessione delle informazioni e, di conseguenza, il *mapping* e il *linking* tra ontologie diverse divengono aspetti essenziali per l'accesso alla conoscenza e per il suo arricchimento, come testimoniato dagli sviluppi sempre maggiori della ricerca in questo ambito (Otero-Cerdeira et al., 2015).

L'esigenza di massimizzare le connessioni tra risorse diverse si deve confrontare con il fatto che ogni ontologia è costruita con criteri differenti, che fanno capo a differenti quadri teorici. In questo contesto l'*instance matching* diventa particolarmente rilevante, poiché consente di collegare risorse senza mappare le entità ontologiche (Castano et al., 2008; Nath et al., 2014).

In questo articolo presentiamo un'ipotesi di collegamento tra due ontologie linguistiche, BabelNet (Navigli and Ponzetto, 2012a) e IMAGACT (Moneglia et al., 2014a), entrambe multimediali,

<sup>1</sup>Lorenzo Gregori ha scritto le sezioni 1 e 3 ed ha sviluppato gli algoritmi di *linking*; Andrea Amelio Ravelli ha scritto le sezioni 2 e 4 ed ha realizzato e valutato i dataset; Alessandro Panunzi ha supervisionato il lavoro e la scrittura dell'articolo. La ricerca è stata condotta nell'ambito del progetto MODELACT, programma Futuro in Ricerca 2012. Project code: RBFR12C6O8 2012-2015.

<sup>2</sup>In questo ambito è particolarmente rilevante l'iniziativa del Linguistic Linked Open Data Cloud (Chiaros et al., 2011), che raccoglie e collega ontologie linguistiche in RDF e ad oggi contiene più di 500 risorse.

multilingui e sfruttabili per task di traduzione e disambiguazione (Moro and Navigli, 2015; Russo et al., 2013; Moneglia, 2014). Il collegamento tra le ontologie avviene attraverso la componente visuale di IMAGACT, ovvero la rappresentazione delle azioni per mezzo di scene prototipiche.

## 2 Risorse

### 2.1 IMAGACT

IMAGACT è un'ontologia dell'azione in cui le entità di riferimento sono identificate attraverso scene (video o animazioni 3D). I verbi che nelle varie lingue si riferiscono allo stesso concetto azionale sono collegati a una stessa scena, che rappresenta prototipicamente tale concetto. La raffigurazione visiva dell'azione veicola l'informazione a livello cross-linguistico: in questo modo si configura come uno strumento di disambiguazione del riferimento azionale, task particolarmente problematico date le differenti strategie di lessicalizzazione dell'azione adottate dalle lingue naturali.

All'interno del sistema di disambiguazione di IMAGACT, le scene prototipiche svolgono la funzione di interlingua. L'organizzazione dell'ontologia non è però regolata da principi lessicografici, ma da criteri di identificazione cognitiva dei concetti azionali categorizzati nelle diverse lingue. Di conseguenza, a partire da una scena è possibile ottenere i corrispettivi lemmi in tutte le lingue presenti in IMAGACT, con la certezza che tali lemmi siano adatti a descrivere l'azione di riferimento, e siano quindi intertraducibili quando applicati al contesto in oggetto (Panunzi et al., 2014).

La risorsa è stata creata a partire dall'analisi di corpora di parlato italiano e inglese attraverso giudizi di categorizzazione espressi da annotatori madrelingua. In questo modo sono stati associati oltre 500 lemmi verbali (per ciascuna lingua) alle stesse 1010 scene prototipiche. I video così ottenuti sono stati in seguito utilizzati per estendere l'ontologia ad altre lingue tramite giudizi di competenza da parte di annotatori madrelingua. Ad oggi, l'interfaccia web di IMAGACT<sup>3</sup> contiene dati completi su quattro lingue (italiano, inglese, spagnolo e cinese mandarino). Il database è in continua espansione (Moneglia et al., 2014b), e attualmente raccoglie dati da 18 lingue con diversi livelli di completezza (Tabella 1).

<sup>3</sup><http://www.imagact.it>

| Lingue     | Verbi BN | Verbi IM |
|------------|----------|----------|
| English    | 57.996   | 1.299    |
| Spanish    | 16.832   | 735      |
| Italian    | 15.590   | 1.100    |
| Portuguese | 11.517   | 792      |
| German     | 5.210    | 992      |
| Chinese    | 4.299    | 1.171    |
| Norwegian  | 2.227    | 107      |
| Danish     | 1.980    | 73       |
| Polish     | 1.910    | 1.145    |
| Hindi      | 342      | 189      |
| Urdu       | 187      | 73       |
| Bangla     | 117      | 120      |
| Serbian    | 91       | 1.145    |
| Sanskrit   | 35       | 198      |
| Oriya      | 6        | 140      |
| Totale     | 118.339  | 9.273    |

Tabella 1: Le 15 lingue comuni di BabelNet (BN) e IMAGACT (IM) con il relativo numero di verbi.

### 2.2 BabelNet

BabelNet<sup>4</sup> è una rete semantica multilingue, un dizionario enciclopedico strutturato in ontologia. Alla base della risorsa vi è la combinazione di due tra le più importanti basi di conoscenza, una linguistica e una enciclopedica, liberamente disponibili online: WordNet e Wikipedia. Si tratta, ad oggi, della più estesa risorsa multilingue per la disambiguazione semantica, giunta alla versione 3.0, che ricopre 271 lingue.

Informazione semantica e informazione enciclopedica sono state raccolte e collegate attraverso un algoritmo di mappatura automatica, al fine di creare un dizionario di concetti ed entità caratterizzato sia da ricchezza informativa, sia da una fitta rete di rapporti semantici a livello ontologico.

Concetti ed entità sono rappresentati per mezzo di BabelSynset (da qui in avanti, BS), estensione del concetto di *synset* utilizzato in WordNet. Un BS corrisponde a un concetto unitario a cui sono collegate le parole che nelle varie lingue vi si riferiscono, corredate da proprietà semantiche, una glossa ed esempi d'uso<sup>5</sup>. Un ulteriore contributo giunge dalla famiglia di risorse collegate a Wikipedia<sup>6</sup>, attraverso cui è stato possibile collegare

<sup>4</sup><http://babelnet.org>

<sup>5</sup>Tali collegamenti si basano anche su relazioni ereditate dai vari Multilingual WordNet(s).

<sup>6</sup>Wiktionary, OmegaWiki (versione allineata dei Wiktionary nelle singole lingue), Wikidata (database document-

ai BS le immagini archiviate in Wikimedia Commons e offrire così tale informazione a supporto della disambiguazione.

### 3 L'esperimento di *linking*

L'esperimento si colloca nello scenario più generale in cui viene realizzato un collegamento tra un'ontologia di dominio aperto, che raccoglie un'ampia serie di concetti eterogenei poco specificati, e una specialistica, dove i concetti di un singolo dominio sono rappresentati in modo più dettagliato (Magnini and Speranza, 2002).

Il nostro esperimento sfrutta i verbi equivalenti in traduzione nelle due risorse per effettuare il *linking* tra le ontologie attraverso le scene prototipali presenti in IMAGACT. Il criterio utilizzato per trovare le scene che possono rappresentare i concetti azionali di BabelNet è quello della maggior corrispondenza tra i verbi collegati allo stesso BS e quelli collegati alle scene di IMAGACT<sup>7</sup>.

Un diverso tentativo di *mapping* su IMAGACT è stato condotto da De Felice et al. (2014) attraverso l'analisi della collegabilità tra i tipi azionali di IMAGACT e i *synset* di WordNet (per l'inglese) e ItalWordNet (per l'italiano). Anche in quell'esperimento è stato sfruttato l'insieme dei verbi comuni come indice di similarità tra gli oggetti delle diverse ontologie. L'ambito di applicazione del nostro esperimento è però diverso, poiché è allargato ad un contesto multilingue: prima di tutto non si tratta di un *mapping* tra concetti delle due ontologie, ma di un *linking* tra istanze di tipo diverso; in secondo luogo non si usa WordNet, ma BabelNet e, coerentemente, il dato di IMAGACT che viene sfruttato non corrisponde ai tipi (che dipendono dai diversi lemmi delle varie lingue) ma alle scene (oggetti di riferimento interlinguistico).

#### 3.1 Il dataset

L'esperimento che presentiamo è stato condotto su un dataset di riferimento annotato manualmente e composto da 25 scene di IMAGACT e 30 BS; per ciascuna coppia  $\langle \text{scena}, \text{BS} \rangle$  è stato giudicato se la scena fosse o meno adatta a rappresentare il BS. Il dataset, comprendente 750 giudizi binari (25 scene per 30 BS), è stato utilizzato per la valutazione degli algoritmi automatici<sup>8</sup>. Per selezionare le sce-

*oriented* che contiene risorse multimediali).

<sup>7</sup>Per il test è stata utilizzata la versione 3.0 di BabelNet; i dati sono stati estratti utilizzando l'API Java (Navigli and Ponzetto, 2012b)

<sup>8</sup>Il dataset è disponibile alla pagina <http://bit.ly/1MtZqB9>

ne sono stati considerati i prototipi azionali della variazione di 7 verbi inglesi (*put*, *move*, *take*, *insert*, *press*, *give* e *strike*), che proiettano 152 BS totali; di questi, ne sono stati scelti 25. La selezione è stata fatta in modo casuale, ma ha tenuto conto del fatto che non tutte le scene sono collegate allo stesso numero di verbi, sia perché le lingue di IMAGACT non sono egualmente rappresentate (vedi tabella 1), sia perché è diverso il numero di verbi utilizzabili per riferirsi alle azioni nelle diverse lingue. Per creare un *test set* che fosse un campionamento rappresentativo, abbiamo cercato di preservare queste differenze quantitative inserendo scene con un diverso numero di verbi collegati (da un minimo di 7 a un massimo di 18) in modo proporzionale all'intero set di scene di IMAGACT. Anche il numero di verbi contenuti in un BS è molto variabile, per cui la loro selezione ha seguito un criterio simile: ognuno dei 30 BS del dataset ha da un minimo di 4 ad un massimo di 51 verbi collegati. Il dataset pubblicato è derivato da un *agreement*: sono stati utilizzati tre annotatori ed il giudizio inserito è quello di almeno due annotatori su tre. L'*inter-rater agreement* riporta una *k* di Fleiss pari a 0,76.

#### 3.2 Gli algoritmi

L'algoritmo base (Algoritmo 1) utilizzato per il *linking* si avvale di una funzione che calcola la vicinanza tra una scena e un BS misurando la frequenza con cui i verbi collegati alla scena di IMAGACT sono legati anche al BS. L'insieme dei candidati è composto da tutti i BS che sono concetti possibili per ogni verbo collegato alla scena.

---

##### Algoritmo 1 Algoritmo base

---

- 1:  $s$ : scena in ingresso
  - 2:  $V$ : set di verbi collegati a  $s$  in IMAGACT
  - 3:  $\text{List}(\text{BabelSynset})$   $LS$ : lista vuota
  - 4: per ogni  $v_i$  in  $V$
  - 5:      $\text{List}(\text{BabelSynset})$   $Syn$  = lista di BS collegati a  $v_i$
  - 6:     aggiungi  $Syn$  a  $LS$
  - 7:  $\text{List}(\text{BabelSynset})$   $FLS$  =  $\text{freqList}(LS)$
  - 8: Collega  $s$  ai primi  $n$  BabelSynset di  $FLS$
- 

La funzione *freqList* calcola la lista di frequenza dei BS in  $LS$  e li ordina dal più frequente al meno frequente.

A partire da questa versione di base dell'algoritmo di *linking*, è stata implementata una versione migliorata che sfrutta la rete semantica di BabelNet, in modo da includere nell'analisi anche i BS semanticamente vicini. Anziché estrarre gli equivalenti in traduzione soltanto dai BS collegati di-

rettamente ai verbi, vengono considerati anche i verbi appartenenti a *synset* collegati al BS principale tramite le relazioni di BabelNet fino a un certo livello di profondità. Per pesare in modo differenziato i BS collegati al verbo abbiamo utilizzato una funzione ricorsiva  $w$ , definita nel modo seguente.

Dato  $S$  l'insieme dei BS,  $s_0 \in S$  collegato direttamente al verbo e  $s', s'' \in S$  collegati tra loro da una relazione  $r \in R$ , definiamo una funzione  $w : S \rightarrow [0, 1]$  tale che  $w(s_0) = freq(s_0)$  e  $w(s'') = w(s') \cdot c \cdot p(r)$ , dove:  $freq$  calcola la frequenza del BS così come riportato nell'algoritmo base;  $R$  è l'insieme delle relazioni tra BS verbali;  $p(rel) : R \rightarrow [0, 1]$  è una funzione che assegna un peso ad ogni relazione  $R$ ;  $c \in [0, 1]$  è un coefficiente di riduzione di peso all'aumentare della distanza dal nodo centrale.

Questa metrica consente di differenziare l'importanza delle diverse relazioni semantiche di BabelNet. Abbiamo infatti verificato che, mentre alcune di esse sono molto rilevanti per il task, altre importano informazione non pertinente e devono quindi essere escluse. La tabella seguente mostra l'elenco delle relazioni tra BS verbali con il relativo valore di rilevanza misurato con *information gain* sul dataset annotato.

| Relazioni in BabelNet | Valore IG |
|-----------------------|-----------|
| 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     |

Tabella 2: Relazioni tra BS verbali.

## 4 Valutazione

### 4.1 Risultati della valutazione

I due algoritmi sono stati eseguiti sulle 25 scene del dataset, quello di base e quello migliorato (considerando soltanto un livello di profondità). È stata quindi verificata l'aderenza dei primi  $n$  BS estratti dagli algoritmi alle scene. La tabella 3 riporta la sintesi dei risultati ottenuti<sup>9</sup>.

Per entrambi gli algoritmi il primo BS candidato per il *linking* è sempre corretto. I risultati

|                     | Alg. base | Alg. migliorato |
|---------------------|-----------|-----------------|
| % corr. ( $n = 1$ ) | 100%      | 100%            |
| % corr. ( $n = 2$ ) | 84%       | 88%             |
| % corr. ( $n = 3$ ) | 76%       | 83%             |

Tabella 3: Percentuale di assegnazioni corrette di scene a BS con i due algoritmi e al variare di  $n$ .

peggiorano progressivamente al crescere di  $n$  e, parallelamente, aumenta anche il divario qualitativo tra i due algoritmi. Inoltre, il dataset annotato è stato utilizzato come *training set* per due ulteriori algoritmi di *machine learning*, uno che considera soltanto *features* relative ai BS collegati in modo diretto, l'altro che include anche *features* dei BS collegati indirettamente attraverso relazioni semantiche. Il risultato<sup>10</sup> riporta rispettivamente 81,16% e 86,95% di assegnamenti corretti di scene al BS. Benché il *test set* sia troppo piccolo per avere una stima precisa sull'efficacia, i risultati sono incoraggianti e compatibili con quelli ottenuti dai due algoritmi (semplice e migliorato) eseguiti sul dataset. La differenza tra le due percentuali mostra chiaramente che l'utilizzo dei BS vicini è significativo per questo task.

### 4.2 Conclusioni

Benché non sia ancora stato fatto un *fine-tuning* dei parametri (per il quale è necessario un *test set* più ampio), i buoni risultati ottenuti da questo esperimento aprono la possibilità di collegare le due ontologie attraverso le scene di IMAGACT, al fine di arricchire entrambe le risorse. Da un lato, i video di IMAGACT potrebbero rappresentare i concetti azionali dei BS; dall'altro, IMAGACT verrebbe arricchita con l'informazione di traduzione presente in BabelNet.

Inoltre, dall'osservazione di Babelfy (Moro et al., 2014), motore di *word sense disambiguation* e *entity linking* derivato da BabelNet, è apparso evidente che l'ipotesi di *linking* qui proposta avrebbe un notevole impatto sull'espressività della rappresentazione visuale delle frasi, con l'associazione di immagini a nomi e di video a verbi.

Infine, è importante notare che sia BabelNet sia IMAGACT sono risorse in espansione: poiché gli algoritmi sfruttano gli equivalenti in traduzione, i risultati potranno essere via via più precisi all'aumentare delle lingue e dei lemmi considerati.

<sup>9</sup>Per i risultati completi consultare <http://bit.ly/1MtZqB9>

<sup>10</sup>L'algoritmo utilizzato è SVM con *kernel* lineare e la valutazione è fatta con *10-folds cross-validation*.

## References

- 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.
- Bernardo Magnini and Manuela Speranza. 2002. Merging Global and Specialized Linguistic Ontologies. In *Proceedings of the Workshop Ontolex-2002 Ontologies and Lexical Knowledge Bases, LREC-2002*, pages 43–48.
- 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, 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.
- Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2:231–244.
- 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. Rodriguez-Martinez, and Alma Gmez-Rodriguez. 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, Churen 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.