

Annotating opposition among verb senses: a crowdsourcing experiment

Anna Feltracco^{1,2}, Elisabetta Jezek², Bernardo Magnini¹, Simone Magnolini^{1,3}

¹Fondazione Bruno Kessler, Povo-Trento

² University of Pavia, Pavia

² University of Brescia, Brescia

{feltracco, magnini, magnolini}@fbk.eu, jezek@unipv.it

Abstract

English. We describe the acquisition, based on crowdsourcing, of opposition relations among Italian verb senses in the T-PAS resource. The annotation suggests the feasibility of a large-scale enrichment.

Italiano. *Descriviamo l'acquisizione, basata su crowdsourcing, di relazioni di opposizione tra sensi di verbi italiani nella risorsa T-PAS. L'annotazione mostra la fattibilità di un arricchimento su larga scala.*

1 Introduction

Several studies have been carried out on the definition of opposition in linguistics, philosophy, cognitive science and psychology. Our notion of opposition is based on lexical semantic studies by Lyons (1977), Cruse (1986; 2002; 2011), and Pustejovsky (2000), as synthesized in Jezek (2015).

The category of opposites can be said to include pairs of terms that contrast each other with respect to one key aspect of their meaning, such that together they exhaust this aspect completely. Examples include the following pairs: *to open / to close*, *to rise / to fall*. Paradoxically, the first step in the process of identifying a relationship of opposition often consists in identifying something that the meanings of the words under examination have in common. A second step is to identify a key aspect in which the two meanings oppose each other¹. Opposites cannot be true simultaneously of the same entity, for example a *price* cannot be

said *to rise* and *to fall* at exactly the same point in time.

It is an open discussion whether opposition is a semantic or a lexical relation (Murphy, 2010; Fellbaum, 1998); what is clear is that the predicate that is considered opposite of another predicate does not activate this relation for all its senses. For example, the Italian verb *abbattere* is considered opposite to *costruire* as far as the former is considered in its sense of *to destroy (a building)*, and the latter in its sense of *to build (a building)*. The opposition relation does not hold if *abbattere* is considered in its sense of *to kill (an animal)*.

Oppositions between verbs senses are poorly encoded in lexical resources. English WordNet 3.1 (Miller et al., 1990) tags oppositions among verb senses using the label *antonymy*; for example, *increase#1* is in antonymy relation with *decrease#1*, *diminish#1*, *lessen#1*, *fall#11*. In VerbOcean (Chklovski and Pantel, 2004), opposition (*antonymy*) is considered a symmetric relation between verbs, which includes several subtypes; the relation is extracted at verb level (not at sense level). FrameNet (Ruppenhofer et al., 2010), on the other hand, has no tag for the opposition relation, although a subset of cases can be traced via the *perspective on* relation. As regards Italian, in MultiWordNet (Pianta et al., 2002) the opposition relation (labeled: *antonymy relation*) is considered a lexical relation and is represented in the currently available version for English, but not for Italian. In SIMPLE (Lenci et al., 2000), the opposition relation (*antonymy*) is considered a relation between word senses and it has been defined for adjectives (e.g., *dead/alive* and *hot/cold*), although the authors specify it can possibly be extended also to other parts of speech. In Senso Comune (Oltremari et al., 2013) the annotation of the opposition relation appears not to be implemented, even if the tag for the relation (*antonimia*) is present.

The experiment described in the paper focuses

¹According to Cruse, opposites indicate the relation in which two terms typically differ along only one dimension of meaning: in respect of all other features they are identical (Cruse, 1986, p.197).

on the annotation of opposition relations among verb senses. We annotate these relations in the lexical resource T-PAS (Jezek et al., 2014), an inventory of typed predicate argument structures for Italian manually acquired from corpora through inspection and annotation of actual uses of the analyzed verbs. The corpus instances associated to each T-PAS represent a rich set of grounded information not available in other resources and facilitate the interpretation of the different senses of the verbs².

We collected data using crowdsourcing, a methodology already used in other NLP tasks, such as Frame Semantic Role annotation (Fossati et al., 2013; Feizabadi and Padó, 2014), Lexical Substitution (Kremer et al., 2014), Contradictory Event Pairs Acquisition (Takabatake et al., 2015).

2 The T-PAS Resource

The T-PAS resource (Jezek et al., 2014) is a repository of Typed Predicate Argument Structures for Italian acquired from corpora by manual clustering of distributional information about Italian verbs. T-PASs are corpus-derived verb patterns with specification of the expected semantic type (ST) for each argument slot, such as `[[Human]] guida [[Vehicle]]`. T-PAS is the first resource for Italian in which semantic selection properties and sense-in-context distinctions of verbal predicates are characterized fully on empirical ground. We discover the most salient T-PASs using a lexicographic procedure called Corpus Pattern Analysis (CPA) (Hanks, 2004), which relies on the analysis of co-occurrence statistics of syntactic slots in concrete examples found in corpora.

The resource consists of three components³:

1. a repository of corpus-derived T-PASs linked to lexical units (verbs);
2. an inventory of about 230 corpus-derived semantic types for nouns (HUMAN, EVENT, BUILDING, etc.), relevant for the disambiguation of the verb in context (see Table 1)
3. a corpus of sentences instantiating the T-PASs⁴.

²The experiment is part of a broader project consisting in enriching the T-PAS resource with the annotation of different types of opposition relation not present in other resources.

³The first release of T-PAS contains 1000 analyzed average polysemy verbs. T-PAS is freely available under a Creative Commons Attribution 3.0 license at tpas.fbk.eu.

⁴The reference corpus is a reduced version of ItWAC (Baroni and Kilgariff, 2006).

Verb: <i>abbattere</i>
▷ T-PAS 1:
<code>[[Human]] abbattere [[Animate]]</code>
▷ Annotated Corpus:
..Kenai, il più giovane, <i>abbatte</i> l' orso..
..un bracconiere <i>abbatteva</i> un coniglio..
...
▷ T-PAS 2:
<code>[[Human Event]] abbattere [[Building]]</code>
▷ ...
▷ T-PAS n°
▷ ...

Table 1: T-PAS Resource Structure.

At present, T-PASs are not linked by any semantic relation. Our experiment extends the resource by adding opposition relations among T-PASs following a pilot experiment described in Feltracco et al. (2015). In the following sections, we illustrate the annotation tasks we elaborated and the new experiments we performed, together with their evaluation.

3 Annotation Tasks

In this section we define the annotation tasks (Section 3.1), how such tasks have been implemented using T-PAS (Section 3.2) and the crowdsourcing platform we used to collect the data (Section 3.3).

3.1 Tasks Definition

In order to annotate the opposition relation among T-PASs, we have set the experiment in two steps. In the first step (Task A) we want to determine if there is an opposition relation between a certain sense of a verb (the *source verb*) and another verb (the *target verb*); in the second step (Task B) we want to identify which sense of the Target Verb holds the opposition relation with the source verb, if identified in Task A.

As for Task A (see an example in Table 2), we showed annotators a pair of sentences: S1 (i.e. Frase 1 in Table 2) is a sentence, extracted from the annotated corpus of T-PAS, that contains the source verb (in bold), while S2 (i.e. Frase 2 in Table 2) is identical to S1, with the exception of the source verb, which is substituted with the target verb (in bold). Annotators were asked to compare the two sentences and choose one among the following options: A1) S2 makes sense and holds an opposition relation with S1, or A2) S2 makes sense but it does not hold an opposition relation with S1, or A3) S2 does not make sense⁵.

⁵Task A is comparable to a Lexical Substitution task. For

Confronta le seguenti frasi.	
Frase 1: L' appello va, pertanto, respinto . (annotated example of T-PAS 1 of the source verb = <i>respingere</i>)	
Frase 2: L' appello va, pertanto, approvato . (Target Verb = <i>approvare</i>)	
Task A: “Diresti che la Frase 2 ha un senso compiuto? Se sì, diresti che c’è una relazione di opposizione tra le Frasi?”	
A1: Frase 2 ha senso e si oppone alla Frase 1 (30.9%)	Inter-Annotator Agreement Ao: 72.4% Fleiss’s coefficient: 0.44
A2: Frase 2 ha senso ma non si oppone alla Frase 1 (5.1%)	
A3: Frase 2 non ha senso (64%)	
Task B: “Leggi le seguenti frasi. In quali frasi approvare ha lo stesso senso della Frase 2?”	
B1: La Commissione approva l’emendamento 2.15 del relatore.	Average Agreement Normalized Ao: 71.7% M. A. Fleiss’s coefficient: 0.32
B2: Gli astronauti hanno approvato l’ uso del TVIS in questa configurazione.	
B3: In ogni caso , non verranno approvati i candidati che abbiano registrato assenze superiori a un terzo del numero complessivo di ore di lezione previste.	
B4: Nessuna delle precedenti	

Table 2: Example and Results for Task A and Task B (Ao: Observed Agreement, M.A.: Macro Average).

If a relation of opposition was identified, we asked annotators to complete Task B. In Task B, annotators had to consider the target verb in S2 and select among a list of sentences containing that verb, those in which the target verb has the "same" meaning as in S2 (Table 2) ⁶.

3.2 Tasks Implementation

Tasks implementation required: (i) the selection of the source verb and target verb, (ii) the extraction of the examples (for S1), and (iii) the substitution of the verb in the examples (for S2).

Verbs Selection. For the annotation of the T-PAS resource, we selected pairs of verbs (source verb and target verb) according to three conditions: (i) both verbs are present in the T-PAS resource; (ii) both verbs appear in the Dizionario dei Sinonimi e dei Contrari - Rizzoli Editore⁷ as lemmas; (iii) the target verb is annotated as *contrary* for the source verb and viceversa in the Dizionario dei Sinonimi e dei Contrari; thus, for each pair source verb A - target verb B, also the pair source verb B - target verb A has been considered. The total number of verb pairs extracted according to these criteria is 436. Since our aim is to annotate opposition among verb

instance, in McCarthy and Navigli (2009) and Kremer et al. (2014), annotators are asked to provide a synonym for a word in a sentence that would not change the meaning of the sentence. In our case we asked annotators to validate the sense of a sentence in which a word is substituted with a supposed opposite.

⁶In other Word Sense Disambiguation (WSD) tasks, e.g. in Mihalcea (2004), annotators are asked to select among a sense inventory. By contrast, we showed non-expert annotators in the crowd the target verb in context, taking advantage of the availability of examples of the verb in the resource.

⁷http://dizionari.corriere.it/dizionario_sinonimi_contrari

senses, we implemented Task A for each of the T-PASs of the source verbs (i.e. for T-PAS 1 of the source verb *abbattere*, for T-PAS 2, for T-PAS 3, ..), for a total of 2263 T-PASs.

Examples Extraction. In order to increase the reliability of the annotation, we extracted up to three examples for each sense of the verbs from the T-PAS resource, according to their availability in the resource (i.e., we extracted up to three examples for the T-PAS 1 of the source verb *abbattere*, up to three examples for the T-PAS 2, ..). We discarded examples annotated as "non regular" such as metonymical uses and, to simplify the task, we selected the shortest examples, composed by at least 5 tokens. The extracted examples for a verb have been used both as the S1 in Task A (when it is the source verb), and as the answers proposed in Task B (when it is the target verb).

Verb Substitution. We generated S2 from S1 substituting the source verb with the target verb automatically conjugated accordingly, using the library: *italian-nlp-library*⁸. The library analyzes only the verb and not the whole sentence and does not manage all the suffixes; to solve this we added some simple rules. This system grants a quick implementation avoiding parsing or deeper analysis of the sentence.

3.3 Crowdfower Platform Settings

For crowdsourcing we used the Crowdfower platform⁹, with the following parameter setting. We

⁸<https://github.com/jacopofar/italian-nlp-library>

⁹<http://www.crowdfower.com>

initially set the payment to 0.04 USD, then to 0.05 USD for each page and the number of sentence pairs for page to 5¹⁰. One out of these 5 pairs was a Test Question (TQ), i.e. a question for which we already know the answer. If an annotator misses many TQs s/he is not permitted to continue the annotation and his/her judgments are rejected: we set the threshold of this accuracy to 71%. We selected the TQs among the total sentence pairs and we annotated them before lunching the tasks. We also set parameters in order to have annotators with Italian Language skills.

4 Results and Discussion

A total of 712 pairs of sentences has been annotated with 3 judgments in almost a month, for a total of 2136 judgments (plus judgments for TQs).

For Task A, the overall inter-annotator agreement (IAA) calculated using *Fleiss's coefficient* (Artstein and Poesio, 2008) is 0.44, with an Observed Agreement (Ao) of 72.4%. Overall, answer A1 was chosen 30.9% of the times, answer A2 5.1% and answer A3 64%.

We observe many cases in which a mismatch between the verb (in any of its meanings) and the new context in which the verb is inserted invalidates the sense of the sentence in its entirety. For instance, in Example 1, where “ridare” is the source verb and “trattenere” the target verb, the relation between the target verb and the direct object argument produces a sentence which has no sense. The pair has been judged as “Frase 2 non ha senso” by the three annotators, since you can “ridare un esame” (“take an exam again”) but not “trattenere un esame” (*“to hold, to keep an exam”).

- (1) S1: Posso **ridare** un esame già sostenuto?
S2: Posso **trattenere** un esame già sostenuto?

Other cases in which the three annotators chose “Frase 2 non ha senso” depend on the relation between the verb and other elements of the sentence. Example 2 shows a case with a coordinative structure between two events: in S1 somebody has been “imprisoned *and* deported”, in S2 somebody has been “released *and* deported”. We believe that annotators judged the two events in S2 as incompatible.

¹⁰In addition to Task A and B, annotators were asked another question concerning the relation among the two verbs. In this paper we are not discussing this further question.

- (2) S1: Era stato **incarcerato** e deportato.
S2: Era stato **liberato** e deportato.

Task B was proposed to annotators only if an opposition had been identified in Task A (i.e. answer A1). Results are calculated for the pairs which collected a minimum of two (out of three) answers A1, for a total of 211 pairs. We calculated the IAA for each sense, considering a match when annotators agree both on selecting and not selecting a sentence (i.e. a sense). The overall average Ao, normalized by the number of annotators, is 71.7%. In addition, we calculated a *Macro Average-Fleiss's coefficient* (Mihalcea et al., 2004), where also the Expected Agreement (Ae) and the Fleiss's coefficient were determined for each pair, and then combined in an overall average. We calculated Ae *a posteriori*, considering the distribution of judgments of annotators, resulting in a *Macro Average-Fleiss's coefficient* of 0.32¹¹.

As regards the crowdsourcing methodology, although the use of examples in place of sense definition simplified the annotation, the tasks were considered rather difficult by many annotators and most of them were discarded for low accuracy in the initial page (which has only TQs), especially for missing TQs for Task B.

5 Conclusion and Further work

In this paper we have presented a crowdsourcing experiment for the annotation of the opposition relation among verb senses in the Italian T-PAS resource.

The annotation experiment has shown the feasibility of collecting opposition relation among Italian verb senses through crowdsourcing. We propose a methodology based on the automatic substitution of a verb with a candidate opposite and show that the IAA obtained using sense examples is comparable with the IAA obtained by other annotations based on sense definitions.

Ongoing work includes further annotation of the opposition relations in T-PAS using crowd answers and a deep examination of the causes which lead to the generation of sentences with no sense.

¹¹These values are similar to the rates reported in other WSD tasks using definitions of senses and not examples; e.g. IAA in Senseval-2 Verb Lexical Sample by expert annotators (Palmer et al., 2006) is 71% and in Senseval-3 by Contributors over the Web (Mihalcea et al., 2004) IAA is 67.3% with a Macro Average-K of 0.35. However in these tasks the IAA was computed somewhat differently (Palmer et al., 2006).

References

- Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.
- Marco Baroni and Adam Kilgarriff. 2006. Large linguistically-processed web corpora for multiple languages. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, pages 87–90. Association for Computational Linguistics.
- Timothy Chklovski and Patrick Pantel. 2004. Verboccean: Mining the web for fine-grained semantic verb relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-04)*, volume 2004, pages 33–40, Barcelona, Spain, July.
- D. Alan Cruse. 1986. *Lexical semantics*. Cambridge University Press.
- D. Alan Cruse. 2002. Paradigmatic relations of exclusion and opposition II: Reversivity. *Lexikologie: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen: Lexicology: An international handbook on the nature and structure of words and vocabularies*, 1:507–510.
- D. Alan Cruse. 2011. *Meaning In Language: An Introduction To Semantics And Pragmatics*. Oxford University Press, USA.
- Parvin Sadat Feizabadi and Sebastian Padó. 2014. Crowdsourcing annotation of non-local semantic roles. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 226–230, Gothenburg, Sweden, April. Association for Computational Linguistics.
- Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.
- Anna Feltracco, Elisabetta Jezeck, and Bernardo Magnini. 2015. Opposition relations among verb frames. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 16–24, Denver, Colorado, June. Association for Computational Linguistics.
- Marco Fossati, Claudio Giuliano, and Sara Tonelli. 2013. Outsourcing FrameNet to the crowd. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 742–747. Association for Computational Linguistics.
- Patrick Hanks. 2004. Corpus pattern analysis. In *Proceedings of the Eleventh EURALEX International Congress, Lorient, France, Université de Bretagne-Sud*.
- Elisabetta Jezeck, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. 2014. T-PAS; A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, Reykjavik, Iceland, May.
- Elisabetta Jezeck. 2015. *The Lexicon. An Introduction*. Oxford: Oxford University Press.
- Gerhard Kremer, Katrin Erk, Sebastian Padó, and Stefan Thater. 2014. What substitutes tell us - analysis of an “all-words” lexical substitution corpus. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 540–549, Gothenburg, Sweden, April. Association for Computational Linguistics.
- Alessandro Lenci, Federica Busa, Nilda Ruimy, Elisabetta Gola, Monica Monachini, Nicoletta Calzolari, and Antonio Zampolli. 2000. Linguistic specifications deliverable d2, Technical report, University of Pisa and Institute of Computational Linguistics of CNR, Pis.
- John Lyons. 1977. *Semantics, Vol. I*. Cambridge: Cambridge.
- Diana McCarthy and Roberto Navigli. 2009. The English lexical substitution task. *Language resources and evaluation*, 43(2):139–159.
- Rada Mihalcea, Timothy Chklovski, and Adam Kilgarriff. 2004. The Senseval-3 English lexical sample task. In *Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 25–28, Barcelona, Spain, July.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to WordNet: An on-line lexical database*. *International journal of lexicography*, 3(4):235–244.
- M. Lynne Murphy. 2010. *Lexical meaning*. Cambridge University Press.
- Alessandro Oltramari, Guido Vetere, Isabella Chiari, Elisabetta Jezeck, Fabio Massimo Zanzotto, Malvina Nissim, and Aldo Gangemi. 2013. Senso Comune: a collaborative knowledge resource for Italian. In *The People’s Web Meets NLP*, pages 45–67. Springer.
- Martha Palmer, Hwee Tou Ng, and Hoa Trang Dang. 2006. Evaluation of WSD systems. In *Word Sense Disambiguation: Algorithms and Applications*, pages 75–106. Springer.
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, volume 152, pages 55–63.

- James Pustejovsky. 2000. Events and the semantics of opposition. In *Events as grammatical objects*, pages 445–482. Stanford, CA: CSLI Publications.
- Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, Christopher R Johnson, and Jan Schefczyk. 2010. FrameNet II: Extended theory and practice. retrieved November 12, 2013.
- Yu Takabatake, Hajime Morita, Daisuke Kawahara, Sadao Kurohashi, Ryuichiro Higashinaka, and Yoshihiro Matsuo. 2015. Classification and acquisition of contradictory event pairs using crowdsourcing. In *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 99–107, Denver, Colorado, June. Association for Computational Linguistics.