Tagging Semantic Types for Verb Argument Positions

Francesca Della Moretta
University of Pavia / Pavia, Italy
`francesca.dellamoretta01`
`@universitadipavia.it`

Anna Feltracco
Fondazione Bruno Kessler / Trento, Italy
University of Pavia / Pavia, Italy
University of Bergamo / Bergamo, Italy
`feltracco@fbk.eu`

Elisabetta Jezek
University of Pavia / Pavia, Italy
`jezek@unipv.it`

Bernardo Magnini
Fondazione Bruno Kessler / Trento, Italy
`magnini@fbk.eu`

Abstract

English. Verb argument positions can be described by the semantic types that characterise the words filling that position. We investigate a number of linguistic issues underlying the tagging of an Italian corpus with the semantic types provided by the T-PAS (Typed Predicate Argument Structure) resource. We report both quantitative data about the tagging and a qualitative analysis of cases of disagreement between two annotators.

Italiano. *Le posizioni argomentali di un verbo possono essere descritte dai tipi semantici che caratterizzano le parole che riempiono quella posizione. Nel contributo affrontiamo alcune problematiche linguistiche sottostanti l'annotazione di un corpus italiano con i tipi semantici usati nella risorsa T-PAS (Typed Predicate Argument Structure). Riportiamo sia dati quantitativi relativi all'annotazione, sia una analisi qualitativa dei casi di disaccordo tra due annotatori.*

1 Introduction

Words that fill a certain verb argument position are characterised for their semantic properties. For instance, the fillers of the object position of the verb "eat" are typically required to share the fact that they are edible objects, like "meat" and "bread". There has been a vast literature in lexical semantics addressing, under different perspectives, this issue, including the notion of selectional preferences (Resnik, 1997) (McCarthy and Carroll, 2003), the notion of prototypical categories (Rosch, 1973), and the notion of lexical sets (Hanks and Jezek, 2008) (Jezek and Hanks, 2010). However, despite the large theoretical interest, there is still a limited amount of empirical evidences (e.g. annotated corpora) that can be used to support linguistic theories. Particularly, for the Italian language, there has been no systematic attempt to annotate a corpus with semantic tagging of verb argument positions

In this paper we assume a corpus-based perspective, and we focus on manually tagging verb argument positions in a corpus with their corresponding semantic classes, selected from those used in the T-PAS resource (Jezek et al., 2014). We make use of an explicit set of semantic categories (i.e., an ontology of Semantic Types), hierarchically organised (e.g. inanimate subsumes food): we are interested in a qualitative analysis, a rather different perspective with respect to recent works that exploit distributional properties of words filling argument positions (Ponti et al., 2016; Ponti et al., 2017). We run a pilot annotation on a corpus of sentences. We aim at investigating how human annotators assign semantic types to argument fillers, and to what extent they agree or disagree.

A mid term goal of this work is the extension of the T-PAS resource with a corpus of annotated sentences aligned with the T-PASs of the verbs (see section 2). This would have a twofold impact: it would allow a corpus based linguistic investigation, and it would provide a unique dataset for training semantic parsers for Italian.

The paper is structured as follows. Section 2 introduces T-PAS and the ontology of semantic types used in the resource. Section 3 describes the annotation task and the guidelines for annotators. Section 4 presents the annotated corpus and the data of the inter-annotator agreement. Finally,

Section 5 discusses the most interesting phenomena that emerged during the annotation exercise.

2 Overview of the T-PAS resource

The T-PAS resource is an inventory of 4241 Typed Predicate Argument Structures (T-PASs) - for example [[Human]] **partecipa a** 'takes part in' [[Event]] - for 1000 average polysemy Italian verbs, acquired from the ItWaC corpus (Baroni and Kilgarriff, 2006) by manual clustering of distributional information about Italian verbs (Jezek et al., 2014), following the Corpus Patterns Analysis (CPA) procedure (Hanks, 2004) (Hanks and Pustejovsky, 2005) which consists in recognising the relevant structures of a verb and identifying the Semantic Types (STs) for their argument slots by generalizing over the lexical sets observed in a sample of 250 concordances. The current list of about 230 semantic types used in the resource (e.g. human, event, location, artifact - henceforth, STs) is corpus derived, that is, STs are the result of manual generalization over the lexical sets found in the argument positions in the concordances, for example in the [[Event]] argument position of *partecipare* we find *gara*, *riunione*, *selezione*, and so forth. Besides the T-PASs and the hierarchically organized list of STs, the resource contains a corpus of sentences that instantiate the different T-PASs for each verb. Each sentence is therefore currently tagged with the number of the T-PAS it instantiates; the tag is located on the verb. No further information is present in the instance except for the T-PAS number.

3 Annotating Semantic Types

The main goal of the annotation effort reported in this paper is to enrich the annotation already present in the examples associated with each T-PAS. Specifically, given a T-PAS of a verb and an example from the corpus, we annotate the lexical items (in the example) generalised by the STs (in the T-PAS).

For instance, Example (1) shows the T-PAS#1 of the verb *vendere* (Eng. 'to sell'), and a sentence associated to it. The task consists in annotating *prodotti tipici* (Eng. 'traditional products') as a lexical item for [[Inanimate]]-obj.

(1) [[Human | Business Enterprise]] **vendere**
[[Inanimate | Animal]]

"[..] il nome di un'associazione brasiliana

che **vendeva** anche prodotti tipici" [1]

We annotate the content word(s) that is the head-noun both in case of the noun-phrases (NP) (e.g. *give a cake*) and in case of prepositional-phrases (PP) (e.g. *give a cake to his little son*). In the case the head-noun is a quantifier, the quantifier is not tagged but the quantified element is (e.g. *to give a piece of cake*).

Notice that more than one token can be annotated, e.g. in the case of multiword expressions such as *prodotti tipici* in Example (1), and more than one item can be tagged for the same argument position, e.g. in case of coordination, such in *[..] che* **vendeva** *anche prodotti tipici e cartoline*" [2].

In the case an argument is not present in the sentence (for instance, when the subject of the verb is unexpressed), we do not signal this lack.

On the other hand, the annotation accounts for the following cases.

Semantic mismatches. Lexical items are annotated according to the T-PAS; however, the annotator can use a different ST, if she/he thinks the one specified in the T-PAS does not apply. For instance, Example (2) reports another instance of T-PAS#1 of *vendere* in which *lavoro* has been annotated as [[Activity]], a ST not selected by the T-PAS#1 of *vendere* in object position (see the T-PAS in Example (1)).

(2) "il lavoro come qualsiasi altra cosa può essere acquistato e **venduto**."[3]

Syntactic mismatches. We account for cases in which the syntactic role of the lexical items does not match with the one proposed in the T-PAS, e.g. in cases of passive forms of verbs, where the subject and prepositional phrase introduced by *da* correspond respectively to the object and the subject of the active construction. In Example (2), *lavoro* is the syntactic subject of the passive clause, and it is generalized by [[Activity]]) in the object position of the T-PAS. In such cases we annotate both the ST of the lexical item and its grammatical relation using the one in the T-PAS.

Pronouns. In case the argument of the verb is realised as a pronoun, we tag the pronoun without assigning a ST. The pronoun is then linked to the noun(s) it refers to, and this noun is actually

---

[1]Eng. '[..] the name of that Brazilian association that **was selling** traditional products'

[2]Eng. '[..] that **was selling** traditional products and postcards'

[3]Eng. 'jobs can be **sold** and bought just like anything.'

tagged with the ST label. In case the pronoun is agglutinated to the verb (i.e. it is found in the same token of the verb, e.g. *venderla*, Eng. 'to sell it'), the part of the token corresponding to the pronoun is specified and, as just specified, the noun is annotated with the ST.

Impersonal constructions. In case of impersonal constructions with an indefinite pronoun, the pronoun is annotated and the ST it refers to is specified: e.g. *In Germania [..] si **vende** a 10 euro al chilo* [4], *si* is annotated with [[Human]].

We annotated the examples in T-PAS using CAT (Content Annotation Tool)[5], a general-purpose text annotation tool (Bartalesi Lenzi et al., 2012).

4 Results of the Pilot Annotation

The pilot annotation consisted in a selection of 3554 sentences extracted from the current version of T-PAS[6] associated to 25 Italian verbs, selected with different levels of polysemy (from a minimum of 2 to a maximum of 10 T-PASs), and argument structure. The average polysemy of the 25 verbs (i.e. number of senses divided by the number of verbs) is 4.08, and for each T-PAS (sense) we have an average of 34.84 annotated sentences.

The annotation was carried out by a master student in linguistics, who was trained on the T-PAS resource, but had no previous experience in annotation. The annotator was able to tag the 3554 sentences in one month.

Table 1 shows the main data of the pilot annotation. Overall, we annotated 5342 argument positions expressed in the 3554 sentences, with an average of 1.5 argument per sentence. Out of the 230 Semantic Types available in the T-PAS ontology, 99 have been selected during the annotation, which means that we used about 40% of the STs contained in the hierarchy.

Data	Total
# Verbs	25
# T-PASs	102
# Examples	3554
# Examples per T-PAS	34.84
# Semantic Types used	99

Table 1: Pilot annotation results.

4.1 Inter-annotator Agreement

In order to assess the reliability of the annotated data, we run an Inter-Annotator Agreement (IAA) test.[7] We asked a second annotator to annotate a sample of 11 T-PASs associated to 3 different verbs (i.e., *pulire*, *vendere* and *sbottonare*). These verbs were chosen because they correspond to about 10% of the annotated sentences. Moreover, we selected them because they present a low or middle degree of polysemy with respect of the group of 25 verbs initially annotated. The second annotator was provided with the task guidelines and a training session was done to solve potential uncertainties in annotation. The second annotator was trained on a selection of corpus instances derived from verb lemmas, which are not included in the evaluation we report here.

Table 2 shows the results of the IAA for each T-PAS. We measured both the agreement on argument annotation, calculated with the Dice's coefficient (Rijsbergen, 1979), and the agreement on ST annotation, calculated as the accuracy (Manning et al., 2008) among the two annotators. As reported in the last row of Table 2, the average agreement is 0.87 for argument annotation, and 0.83 for ST annotation.

T-PAS	Argument *Dice's value*	ST *Accuracy*
Pulire, T-PAS#1	0.83	0.74
Pulire, T-PAS#2	1	1
Sbottonare, T-PAS#1	0.94	0.89
Sbottonare, T-PAS#2	0.95	0.98
Sbottonare, T-PAS#3	1	1
Sbottonare, T-PAS#4	0.88	0.90
Vendere, T-PAS#1	0.87	0.81
Vendere, T-PAS#2	0.33	0.5
Vendere, T-PAS#3	0.8	1
Vendere, T-PAS#4	1	1
Vendere, T-PAS#5	1	1
Overall average	0.87	0.83

Table 2: Inter Annotator Agreement.

A special case is *vendere T-PAS#2*, which shows the lowest score for both argument and STs annotation. The annotation task allowed annotators to discard sentences which according to their opinion did not fit the sense of the T-PAS taken into consideration. *Vendere T-PAS#2* has only a few corpus instances, which were mostly discarded or

---

[4]Eng. 'In Germany, they sell it at 10 euro per kilo'.
[5]https://dh.fbk.eu/resources/cat-content-annotation-tool
[6]http://tpas.fbk.eu

[7]Cinková et al. (2012) held an IAA on pattern-identification using the CPA procedure in 30 English verbs.

tagged differently by the two annotators, causing low agreement in the results for this T-PAS.

5 Discussion

This Section discusses the most interesting phenomena that emerged during the annotation exercise, particularly in light of the Inter-annotator Agreement.

5.1 Discussion: Argument Tagging

In this paragraph, we focus on the disagreements we found in argument tagging. The annotation task was difficult because the annotators had to identify the semantic structure of the verbs, using syntactic criteria to distinguish whether a lexical element was an argument or not.

Annotating pronouns was also a very demanding process since it implies the identification of co-reference chains. Differences in argument annotation between the two annotators, that impact the arguments Dice score, lie mainly in the annotation of pronouns and in the identification of co-referents. One annotator usually tends to annotate all the pronouns contained in an utterance whereas the other tags only the pronoun which is an argument of the verb taken into consideration. In addition, one usually does not identify co-referents which are lexically realised at great distance of words from the tagged verb, whereas the other sometimes annotates co-referents even if the argument has already been identified. There are also differences concerning the extension of annotation e.g. one interpreted *prodotti tipici* as multiword expression and the other did not. Overall, we obtained good agreement results, although some disagreements still remain even if we tried to reduce potential differences in annotation treating as many cases as possible in the guidelines.

5.2 Discussion: Semantic Type Tagging

The main goal of this section is to analyse the results of IAA on ST selection. Annotators used approximately 40 STs even though their expected number (according to the T-PAS resource) was 11. Table 3 represents the ST usage in the IAA experiment for each T-PAS.

Annotators used approximately the expected number of semantic types with some T-PASs, while with others they used many more. To a higher number of STs employed corresponds a lower ST accuracy score (see Table 1), more

T-PAS	ST Expected *according to the T-PAS*	ST used *A+B*
Pulire, T-PAS#1	4	23
Pulire, T-PAS#2	3	4
Sbottonare, T-PAS#1	2	6
Sbottonare, T-PAS#2	2	4
Sbottonare, T-PAS#3	1	1
Sbottonare, T-PAS#4	1	4
Vendere, T-PAS#1	4	23
Vendere, T-PAS#2	2	3
Vendere, T-PAS#3	3	3
Vendere, T-PAS#4	1	1
Vendere, T-PAS#5	1	1

Table 3: Expected and used STs in the IAA test.

specifically this correlation is shown by *pulire* T-PAS#1, *sbottonare* T-PAS#1,#4, *vendere* T-PAS#1. There are a number of reasons that justify this STs usage. In some cases one annotator tends to tag the entity denoted by single lexical items instead of the generalisations made by the T-PASs. This causes a sentence specific annotation that employs STs that are end nodes in the hierarchy, which do not correspond to the ones in the reference T-PAS. As future work, we plan to develop a methodology to normalize the STs to the appropriate level of abstraction.

There are also linguistic reasons that intervene in the assignment of different STs to the same lexical element. Annotators captured repeatedly the phenomenon known as *inherent polysemy* by tagging the same lexical elements in two totally different ways. An inherent polysemous noun denotes, depending on the context, a single aspect of an entity which is inherently complex, i.e. that can be described simultaneously by more than one ST (see (Jezek, 2016) and references therein). An example is provided by the nouns that denote countries that in our annotation exercise have been tagged as [[Business Enterprise]], [[Institution]] or [[Area]], pointing out their complex nature of territorial, politic and economic entity. In some cases annotators have privileged different semantic components in the ST annotation process. This is due to the context in which the words are embedded, that determines certain interpretations instead of others. However, sometimes the compositionality principle does not strictly define the meaning of an utterance. Hence some lexical items remain underspecified so that they can receive more than one ST at once.

For instance in example (3) one annotator tagged *lente* as [[Artifact]] highlighting its nature

of manufactured object, whereas the other has annotated the lexical item as [[Physical Object Part]] focusing on its nature of constituent element of a bigger object.

(3) "Giles **pulisce** una lente dei suoi occhiali."[8]

Moreover, there are differences is ST assignment caused by regular polysemy (Apresjan, 1974), systematic alternation of meaning that apply to classes of words (Jezek, 2016). IAA results reveal regular polysemy patterns for nouns.

6 Conclusions

We performed a pilot experiment to tag the arguments of verbs, as recorded in the T-PAS resource, with their associated semantic type. We obtained good result in the annotation. By analyzing the cases of inter annotator disagreement, we were able to identify phenomena which lie at the core of such disagreements, such as the presence of inherent polysemous words. Ongoing work includes spelling out the rules for polysemous words tagging more clearly in the guidelines.

References

Iurii Derenikovich Apresjan. 1974. Regular polysemy. *Linguistics*, 32.

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.

Valentina Bartalesi Lenzi, Giovanni Moretti, and Rachele Sprugnoli. 2012. Cat: the celct annotation tool. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC '12)*, pages 333–338.

Silvie Cinková, Martin Holub, Adam Rambousek, and Lenka Smejkalová. 2012. A database of semantic clusters of verb usages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC '12)*, pages 3176–3183.

Patrick Hanks and Elisabetta Jezek. 2008. Shimmering lexical sets. In *Proceedings of the XIII EURALEX International Congress*, pages 391–402.

Patrick Hanks and James Pustejovsky. 2005. A pattern dictionary for natural language processing. *Revue française de linguistique appliquée*, 10(2):63–82.

Patrick Hanks. 2004. Corpus pattern analysis. In *Proceedings of the Eleventh EURALEX International Congress*.

Elisabetta Jezek and Patrick Hanks. 2010. What lexical sets tell us about conceptual categories. *Lexis*, 4(7):22.

Elisabetta Jezek, Bernardo Magnini, Anna Feltracco, Alessia Bianchini, and Octavian Popescu. 2014. T-PAS: a resource of corpus-derived types predicate-argument structures for linguistic analysis and semantic processing. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.

Elisabetta Jezek. 2016. *The lexicon: an introduction.* Oxford University Press.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval.* Cambridge University Press, New York, NY, USA.

Diana McCarthy and John Carroll. 2003. Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, 29(4):639–654.

Edoardo Maria Ponti, Elisabetta Jezek, and Bernardo Magnini. 2016. Grounding the lexical sets of causative-inchoative verbs with word embedding. In *Proceedings of the Second Italian Conference on Computational Linguistic (CLiC-it 2016)*.

Edoardo Maria Ponti, Elisabetta Jezek, and Bernardo Magnini. 2017. Distributed representations of lexical sets and prototypes in causal alternation verbs. *Italian Journal of Computational Linguistics*, to appear.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How*, pages 52–57.

CJ van Rijsbergen. 1979. Information retrieval. 1979.

Eleanor H Rosch. 1973. Natural categories. *Cognitive psychology*, 4(3):328–350.

---

[8]Eng.'Giles **cleans** a lens of his glasses'