Designing a Methodology for Semantic Type Tagging of Argument Positions

Francesca Della Moretta*
Università di Pavia

Anna Feltracco**
Fondazione Bruno Kessler
Università di Pavia

Elisabetta Jezek †
Università di Pavia

Bernardo Magnini ‡
Fondazione Bruno Kessler

A verb argument position can be described by the semantic type that characterizes 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. Our main interest is to evaluate whether our annotation methodology can be employed effectively for the extension of the annotation of the corpus associated with the resource. In order to achieve this goal we compare quantitative data about the tagging and qualitative data derived from the Inter-Annotator Agreement.

1. Introduction

Semantic properties of arguments have been explored under different perspectives by a considerable number of scholars. Most of the theoretical positions share the core tenet that the fillers of a certain argument position are characterized by a set of common semantic features/constraints. 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". This issue has been addressed in a variety of ways, among them 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). Lexical sets are lists of words that regularly occur in particular argument positions in relation to a certain verb pattern (Hanks 2004). It has been shown that these sets contribute greatly to the specification of verb meaning and that they can be generalized by Semantic Types (Pustejovsky 1995). However, despite the large theoretical interest, there is still a limited amount of empirical evidence (e.g. annotated corpora) that can be used to support linguistic theories in this respect. Particularly, for the Italian language, there has been no systematic attempt to annotate a corpus with semantic tagging of verb argument positions.

* Dept. of Humanities - Corso Strada Nuova, 65 - 27100 Pavia, Italy.
E-mail: francesca.dellamoretta01@universitadipavia.it.
** Fondazione Bruno Kessler - Via Sommarive, 18, 38123 Povo TN, Italy.
Dept. of Humanities - Corso Strada Nuova, 65 - 27100 Pavia, Italy.
E-mail: anna.feltracco@gmail.com
† Dept. of Humanities - Corso Strada Nuova, 65 - 27100 Pavia, Italy. E-mail: jezek@unipv.it
‡ Fondazione Bruno Kessler - Via Sommarive, 18, 38123 Povo TN, Italy. E-mail: magnini@fbk.eu

In this paper we assume a corpus-based perspective, and we focus on manually tagging verb argument positions in a corpus with a set of semantic categories (i.e., Semantic Types) hierarchically organized. We are interested mainly in a qualitative analysis, a rather different perspective with respect to recent works that exploit distributional properties of words filling argument positions (Ponti, Jezek, and Magnini 2016, 2017). We run a pilot annotation on a corpus of sentences previously annotated with verb patterns from the T-PAS resource (Jezek et al. 2014), a repository of verb argument structures for Italian (Section 2). 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 design of an annotation task that can be used effectively for the extension of the T-PAS resource with a corpus of annotated sentences aligned with the structures identified for each verb in the resource. A previous attempt of this kind was carried out by Jezek and Frontini (2010); they pointed out for the first time the potential benefits that could derive from the annotation of the corpus instances associated to the different verb argument structures registered in T-PASs.

The extension of the T-PAS resource annotation would have a twofold impact: it would allow a corpus based linguistic investigation of the linguistic phenomena that affect argument selection, and it would provide a unique dataset for training semantic parsers for Italian.

The paper is structured as follows. Section 2 provides information about the T-PAS resource 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 from both a syntactic and a semantic point of view. Section 6 draws the conclusions and provides hints for further work.

2. Overview of the T-PAS resource

The T-PAS resource (Jezek et al. 2014) is an inventory of 4241 Typed Predicate Argument Structures (T-PASs) for Italian. An example of one of these *T-PASs* for the verb *divorare* (Eng. "devour") is given in Example 1.

Example 1
T-PAS#2 of the verb *divorare* (Eng. 'to devour')
[[Human]-subj] *divorare* [[Document]-obj]
example: "La ragazza divorò il romanzo." (Eng. 'The girl devoured the novel.')

Currently, the resource includes typed predicate-argument structures for 1000 average polysemy Italian verbs extracted randomly from the fundamental lemmas of Sabatini Coletti 2008 according to the following proportions: 10% 2-sense verbs, 60% 3-5-sense verbs, 30% 6-11-sense (Jezek et al. 2014). Each structure is acquired from the ItWaC corpus (Baroni and Kilgarriff 2006) by manual clustering of distributional information, following the Corpus Patterns Analysis (CPA) procedure (Hanks 2004; Hanks and Pustejovsky 2005), which consists in recognizing the relevant structures of a verb and identifying the Semantic Types for their argument slots by generalizing over the lexical sets observed in a sample of about 250 concordances. Each argument structure corresponds to a sense of the verb: this sense is expressed in the form of a sense description (called *implicature*) linked to the typing constrains of the structure.

The current list of Semantic Types used in the resource (e.g. human, event, location, artifact) is of 230 Types. In Example 1 above, the Semantic Types [[Human]] and [[Document]][1] are specified for the subject and object position respectively: for [[Human]] as subject of *divorare* we can find a lexical set including words like "ragazza", "io", etc. (Eng. *girl, I, etc.*), for [[Document]] in object position we find "documento", "libro", "romanzo", etc. (Eng. *document, book, novel, etc.*).

The list of Semantic Types is corpus derived, that is, Semantic Types are the result of manual generalization over the lexical sets found in the argument positions in the concordances. In other words, they are verb-motivated. Jezek et al. (2016) mentioned the example of the Semantic Type [[Horse]] which is included in the type list because some Italian verbs select this specific category (e.g. *cavalcare* 'to ride', *ferrare* 'to shoe', *sellare* 'to saddle').

The list is hierarchically organized along the "IS A KIND OF" relation (class subsumption). Figure 1 reports a section of the taxonomy in which [[Flying Vehicle]] is a type of [[Vehicle]] which is a type of [[Artifact]], and so forth.

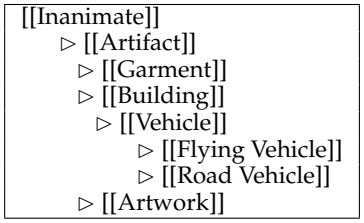

Figure 1
Section of the T-PAS's Semantic Type Hierarchy.

Besides the T-PASs and the hierarchically organized list of Semantic Types, the resource contains a corpus of sentences that instantiate the different T-PASs for each verb. Each sentence is tagged with the number of the T-PAS it instantiates. No further information is currently present in the corpus instances except for the T-PAS number.

Figure 2 reports a schematic representation of the three components of the T-PAS resource: the list of the predicate-argument structures; the inventory of the Semantic Types; the list of corpus sentences.

3. Annotating Semantic Types

The main goal of the annotation exercise we are proposing in this contribution is to identify, in the corpus instances associated to each T-PAS, the lexical elements that fill the argument positions, and tag them with the Semantic Types of the corresponding T-PAS. More in details, given a T-PAS and a sentence in the corpus tagged with that T-PAS, the task consists in:

1. *Argument Tagging*: identifying in the sentence the lexical elements that are arguments of the verb and tag them.
2. *Semantic Type Tagging*: assigning a Semantic Type to the arguments, making reference to those specified in the corresponding T-PAS.

1 Semantic types are indicated with uppercase and between square brackets.

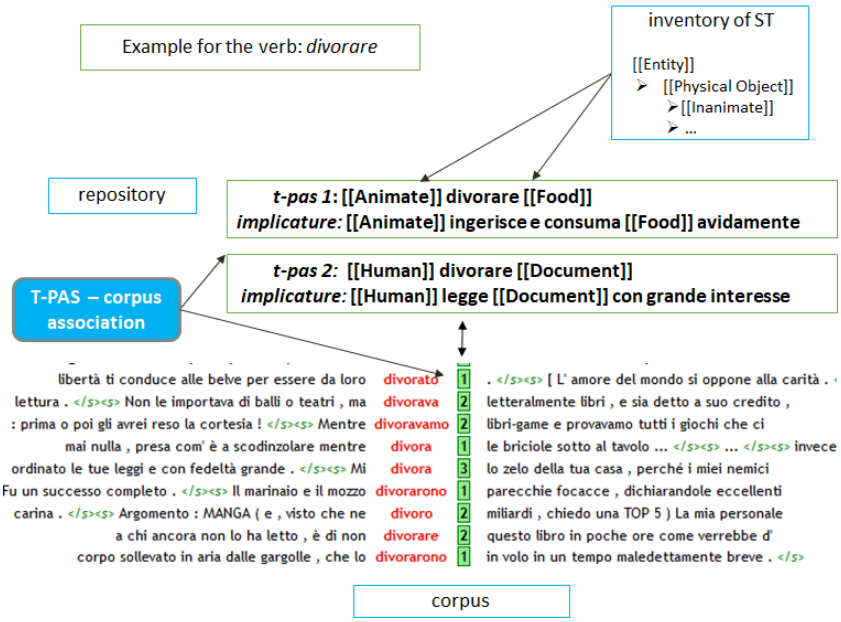

Figure 2
Example of the components of the T-PAS resource: corpus, inventory and repository.

Example 2 shows the T-PAS#1 of the verb *vendere* (Eng. 'to sell'), and a corpus sentence associated to it. The task consists in identifying *prodotti tipici* (Eng. 'traditional products') as object of the verb and tag it as lexical item for [[Inanimate]-obj].

Example 2
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise[2]]-subj] **vendere** [[Inanimate | Animal] -obj]
"[..] un'associazione brasiliana che **vendeva** anche prodotti tipici."
(Eng. '[..] a Brazilian association that **was selling** traditional products')

The task necessitates clear annotation guidelines to cope with the different realizations that a Semantic Type may have, such as lemmas, multiword expressions, pronouns. In general, we annotate content words, that is, the head-noun, both in case of the noun-phrases (NPs) (e.g. 'give a cake') and in case of prepositional phrases (PPs)(e.g. 'give a cake to his little son'). In 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'), since it carries the semantic meaning of the phrase.

We designed a way to tag multiword expressions by allowing annotators to mark more than one lexical item at a time. In Example 2, the multiword expression *prodotti tipici* is annotated as an argument [[Inanimate]-obj] following this methodology.

The annotation accounts also for the following cases.

2 In the T-PASs, | marks the alternation between two Semantic Types in one argument position.

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 the prepositional phrase introduced by *da* (Eng. 'by') correspond respectively to the object and the subject of the active construction which is recorded by the T-PAS. In Example 3, *velivoli* (Eng.'planes') is the syntactic subject of the passive clause, and it is generalized by [[Inanimate]] in the object position of the T-PAS. In such cases, we annotate both the Semantic Type of the lexical item and its grammatical relation (e.g. subj, obj, and so on) using the one specified in the T-PAS. In Example 3 *velivoli* has been annotated as [[Inanimate]-obj].

Example 3
T-PAS#1 of the verb *abbattere*
[[Human | Event]-subj] **abbattere** [[Plant | Inanimate]-obj]
"Quella sera vennero **abbattuti** due velivoli, uno dei quali era civile."
Eng. 'That night two planes were **shut down**, one of those used to be civilian'

Pronouns. If the argument of the verb is realized as a pronoun, we tag the pronoun without assigning a Semantic Type. The pronoun is then linked to the noun(s) it refers to, so that it is possible to identify which Semantic Type label generalizes both the lexical item and the pronoun. For instance, in Example 4 the relative pronoun *che* (Eng. 'that') refers to *banane* (Eng. 'bananas'); first, we link the pronoun to its nominal co-referent, then we tag *banane* as [[Inanimate]-obj]. In this way, it is possible to assign to both lexical items the Semantic Type [[Inanimate]].

Example 4
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"Da noi vi era il monopolio delle banane somale che venivano **vendute** intorno alle 350 lire."
Eng. 'Where we lived there was a monopoly on somalian bananas that were **sold** at around 350 lire.'

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 portion of the token corresponding to the pronoun is tagged, and linked to the noun(s) it co-refers to, in order assign the correct Semantic Type.

Impersonal constructions. In case of impersonal constructions with an indefinite pronoun, we annotate the pronoun itself with the adequate Semantic Type. In Example 5, the token *si* is annotated with [[Human]]:

Example 5
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"In Germania [..] si **vende** a 10 euro al chilo."
Eng. 'In Germany [..] they **sell** it at 10 euro per kilo'.

Coordinated Arguments. An argument position can be filled with more than one item at the same time; these tokens are usually separated by a comma or by the coordinating conjunction *and*; in this case, we tag all the coordinated tokens that fill the same argu-

ment position. In Example 6, both *banane* and *fiori* are tagged as arguments [[Physical Object]-obj].

Example 6
T-PAS#2 of the verb *comprare* (Eng. 'to buy')
[[Human]-subj] **comprare** [[Physical Object]-obj]
"Vi hanno mai detto di **comprare** banane e fiori del commercio equo invece di quelli a buon mercato?"
Eng. 'Haven't they told you to **buy** fair trade bananas and flowers instead of the cheap ones?'

In 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. In Example 6, for example, we have not signalled the absence of the argument [[Human]-subj].

Finally, as regards Semantic Type tagging, we allow annotators to select a Semantic Type other than the one proposed in the particular T-PAS which is being annotated over the corpus, in case it is found that the proposed Semantic Type is not fully adequate to generalize over the lexical items in the corpus.

We annotated the examples in T-PAS using CAT (Content Annotation Tool)[3], a general-purpose text annotation tool (Bartalesi Lenzi, Moretti, and Sprugnoli 2012). Figure 3 shows the CAT tool interface and the annotation of the T-PAS#1 of the verb *vendere*. Annotators can select from a drop-down menu the label to assign to a token; particularly, when the ARGUMENT label is assigned to the token *fucile* 'rifle', a list of Semantic Types appears and the Semantic Type [[Inanimate]] can be selected.

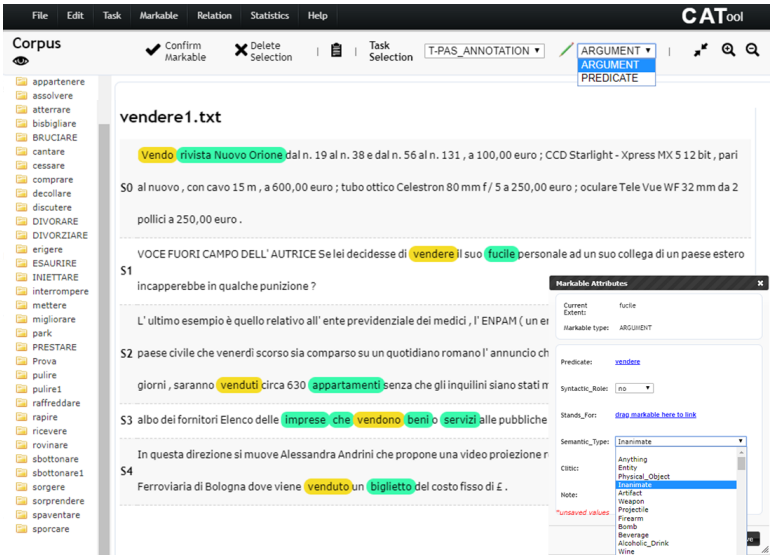

Figure 3
CAT screenshot for the T-PAS#1 of the verb vendere.

3 https://dh.fbk.eu/resources/cat-content-annotation-tool.

4. Results of the Pilot Annotation and of the Inter-Annotator Agreement

We run a pilot annotation on a selection of 3554 sentences extracted from the current version of T-PAS[4] associated to 25 Italian verbs: *abbattere, abbottonare, acquistare, appartenere, assolvere, atterrare, bisbigliare, comprare, decollare, sbottonare, sporcare, tramontare, mettere, urlare, ricevere, sorgere, vincere, pulire, raffreddare, erigere, vendere, migliorare, cessare, interrompere, rapire*.[5] These verbs show different degrees of polysemy in T-PAS: from a minimum of 2 to a maximum of 10 T-PASs. The average polysemy of the 25 verbs (i.e. number of T-PAS divided by the number of verbs) is 4.08, and for each T-PAS we have an average of 34.84 annotated sentences. The annotation was carried out by a master student in linguistics, first author of the current paper, who was trained on the T-PAS resource, but had no previous experience in annotation. It took the annotator about 23 days to tag 3554 sentences. This means that the annotator tagged about 150 sentences a day. Table 1 shows the main data of the annotation. Overall, we annotated 5342 argument positions expressed in the 3554 sentences, with an average of 1.5 argument per sentence.

Table 1
Pilot annotation results.

Data	Total
#Verbs	25
#T-PASs	102
# Examples	3554
# Examples per T-PAS	34.84

In order to assess the reliability of the annotated data, we ran an Inter-Annotator Agreement (IAA) test.[6] 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 their associated examples are 366, which is about 10% of the 3554 annotated sentences. Table 2 summarizes the main data of the IAA test.

Table 2
Results of IAA annotation.

Data	Total
#Verbs	3
#T-PASs	11
# Examples	366
# Examples per T-PAS	33.27

The verbs selected for the IAA test present a lower degree of polysemy with respect to the group of 25 verbs initially annotated. The second annotator was provided with

4 http://tpas.fbk.eu.
5 Eng. 'demolish', 'button up', 'purchase', 'belong', 'absolve', 'land',' whisper', 'buy', 'take off', 'unbutton', 'dirt', 'set', 'put', 'shout', 'receive', 'raise', 'win', 'clean', 'cool', 'erect', 'sell', 'improve', 'cease', 'interrupt', 'kidnap'.
6 A similar investigation was conducted for English by Cinková et al. (2012). The authors held an IAA test on pattern-identification using the CPA procedure for 30 English verbs.

the task guidelines, and a training session was done to solve potential uncertainties in annotation. This training was carried out on a selection of corpus sentences that are not included in the evaluation we report here.[7]

Table 3 shows the results of the IAA for each T-PAS. We measured both the agreement on the *Argument Tagging*, calculated with the Dice's coefficient (Rijsbergen 1979), and the agreement on the *Semantic Type Tagging*, calculated as the accuracy (Manning, Raghavan, and Schütze 2008) among the two annotators. As reported in the last row of Table 3 the average agreement is 0.87 for the *Argument Tagging*, and 0.83 for the *Semantic Type Tagging*.

Table 3
Inter-Annotator Agreement scores.

T-PAS	**Argument Tagging** *Dice's value*	**Semantic Type Tagging** *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

As can be seen from the table, *vendere T-PAS#2* shows the lowest score for both *Argument Tagging* and *Semantic Type Tagging*. This T-PAS presents only few corpus sentences; as a result, the agreement scores are highly affected even by a single case of disagreement between annotators.

5. Discussion

As we have already pointed out, we are interested in defining a methodology that can be employed successfully for the extension of the T-PAS resource, particularly for the annotation of the corpus with Semantic Types. In order to evaluate our annotation exercise we performed a qualitative analysis of the Inter-Annotator Agreement data to identify the most interesting phenomena that not only affected agreement scores, but also influenced the degree of difficulty associated with the task. In the following, we firstly address issues related to the *Argument Tagging*, and secondly issues related to the *Semantic Type Tagging*.

7 The sentences used for the training were chosen on the basis of different criteria to provide the second annotator with a variety of syntactic structures and linguistic phenomena described in the guidelines.

5.1 Discussion: Argument Tagging

We obtained good agreement results for *Argument Tagging*, the scores range between 0.8 and 1 (leaving out *vendere T-PAS#2*) with an average score of 0.87. This subtask is syntactic in its nature; in fact most of the times it consists in identifying the head-nouns of NPs and PPs. Annotators did not experience difficulties in tagging arguments located close to the main verb, and it was also easy for them to tag both impersonal and passive constructions.

Pronoun annotation was assessed as a demanding process, since it implies the identification of co-reference chains.

Differences in argument tagging between the two annotators, that impact the arguments Dice score, involve argument extension and argument identification, on which we report below.

5.1.1 Argument Extension

Argument extension can be defined as the number of tokens that compose an argument. We noticed that the argument extension varied in the annotated data even though annotators identified the same entities as participants to the event. In particular, the major source of disagreements in this respect was the tagging of multi-word expressions. For example, the multi-word expression *prodotti tipici* in Example 2 has been tagged entirely by one annotator and partially by the other: i.e. only *prodotti* has been annotated by the latter.

This fact caused some disagreements that influence the Dice's score (see second column of Table 3), which was calculated taking into consideration the number of tokens identified by the annotators for each argument. This unsystematic annotation can be avoided by introducing clearer criteria in the guidelines, to distinguish between which tokens can be considered part of a multi-word expression, and which cannot.

5.1.2 Argument Identification

There were some disagreements in argument identification. Results show that annotation of pronouns and identification of co-reference chains are the most difficult areas.

Regarding the annotation of pronouns, we noticed that one annotator tags only the pronoun that is an argument of the verb taken into consideration, accordingly to the guidelines, while the other usually tends to annotate all the pronouns that co-refer to the same concept, even if they depend from a predicate that is not the one under consideration. This is what happens in Example 7, where instead of tagging only the clitic *li* of *venderli* as an argument, one annotator also tags the pronoun that depends from the verb *disegnavo* and co-refers to it.

Example 7
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"Pensavo i modelli, li disegnavo, andavo a **venderli** per tutta Italia"
Eng. 'I used to design the new clothing, draw them and then **sell** them all over Italy.'

As regards co-reference chains, one of the annotators usually does not identify co-referents which are lexically realized at long distance of tokens from the tagged verb. For instance, in Example 8 the noun *baccalà* was not tagged as [[Animate=Food]-obj] by one annotator because located at the end of the sentence, far from the pronouns it co-refers to.

Example 8
T-PAS#1 of the verb *pulire*
[[Human]-subj] **pulire** [[Body Part | Artifact | Animate = Food[8]]-obj]
"Poi **pulitelo**, togliendo la pelle e le lische, tagliate a pezzi sui 50 g l'uno, asciugate, infarinate, friggete in olio bollente questi pezzi di baccalà."
Eng. 'After that, **clean** it, by removing the skin and fish bone, slice it into pieces of about 50g each, dry, flour and fry in boiling oil these pieces of codfish.'

Both these issues are brought about mainly by a misunderstanding of the guidelines and can be solved improving the training session of the annotators and further clarifying in the guidelines how to deal with such cases.

5.2 Discussion: Semantic Type Tagging

As presented in Section 3, *Semantic Type Tagging* is the second subtask that annotators accomplished during the annotation exercise. In this section we analyze the data provided by the IAA according to this perspective. Annotators used approximately 40 Semantic Types, even though their expected number (according to the T-PASs being annotated) was 11. Specifically, they used 11 Semantic Types in Subject position and 34 in Object position, whereas the expected ones were respectively 7 and 5. Table 4 summarizes this information.

Table 4
STs expected and used according to each argument position.

Argument Position	Expected Semantic Type	Semantic Types Used
Subject	7	11
Object	5	34

Table 5 represents the Semantic Type usage in the IAA experiment for each T-PAS.

From Table 5, it can be noticed that annotators used approximately the expected number of Semantic Types for some T-PASs, while with others, they used many more. To a higher number of Semantic Types employed corresponds a lower Semantic Type accuracy score (see Table 3); this correlation is evident especially with *pulire* T-PAS#1, *sbottonare* T-PAS#1, and *vendere* T-PAS#1.

Additional results can be found in Table 6, where the employed Semantic Types for each T-PAS in any of its argument positions are indicated.[9] Each Semantic Type filling an argument position of the T-PASs was found and tagged during the annotation exercise, therefore all the expected Semantic Types are recorded in the annotated data.[10]

It has already been said that accuracy scores (see Table 3) tend to drop as the number of Semantic Types employed increases; this can be further captured in the forth column

8 Note that Food is intended here as a role assumed by the Semantic Type. The corpus examples that motivated the inclusion of [[Animal=Food]] in the T-PAS contained lexical items such as *baccalà* 'codfish', *pesce* 'fish' etc. in object position. It can be questioned whether this annotation is too specific for the T-PAS in question; however, we used it as it is currently encoded in the resource.

9 It is important to clarify that each argument position has been tagged by annotators by assigning only one Semantic Type.

10 As explained in section 3, annotators were allowed to use the Semantic Types contained in the ontology if they thought the ones in the T-PAS did not satisfy the features of argument slot fillers.

Table 5
STs expected and used in the IAA test.

T-PAS	**Expected ST** *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

of Table 6, where the list of Semantic Types used by each annotator is provided (the subscripted letters located next to each Semantic Type identify the annotators: A is one annotator, B is the other). In particular, a large number of Semantic Types is present for the object position of some T-PASs, namely *pulire* T-PAS#1, *sbottonare* T-PAS#1, and *vendere* T-PAS#1.

Table 6 also shows that each transitive verb sense in our data has the type [[Human]] or the [[Human | Institution]] type alternation in the subject position. It is the object position, then, that helps the most to disambiguate between different verb senses.

A possible interpretation of these data is that annotators used some non-expected Semantic Types because the expected ones specified in the T-PAS seemed somewhat unsatisfactory. This is the case of Example 9 where *bottone* (Eng. 'button') has been tagged as [[Physical Object Part]] by annotator A and as [[Artifact]] by annotator B; this behaviour has been repeated for all the instances of *bottone* in the object position for T-PAS#2 of the verb *sbottonare*.

Example 9
T-PAS#2 of the verb *sbottonare*
[[Human]-subj] **sbottonarsi** [[Garment]-obj]
"Il controllore si **sbottonò** il primo bottone della camicia [..]"
Eng. 'The ticket inspector **unbottoned** the first botton of his shirt[..]'

These annotations show that annotators had to find another Semantic Type to describe *bottone* because [[Garment]] was considered not adequate. As specified, they select: [[Physical Object Part]], which emphasizes the nature of *bottone* as component of a piece of cloth, and [[Artifact]], which emphasizes the nature of *bottone* as a processed element. A systematic analysis of these cases provides us with useful information to enrich the T-PASs with new Semantic Types in the corresponding argument positions.

In other cases, it appears that annotators felt that the expected Semantic Type was too general and therefore gave a poor description of the annotated token(s). In Example 10 one annotator employed the Semantic Type [[Gas]] to tag the lexical item *gas*, which

Table 6
Semantic types used and expected in the IAA data.

T-PAS	Argument Position	Expected ST	ST Used A + B
Pulire T-PAS1	Subj	**[[Human]]**	**[[Human]]**$_{AB}$, [[Institution]]$_{A}$ [[Human Group]]$_{A}$ [[Business Enterprise]]$_{A}$ [[Artifact]]$_{A}$
	Obj	**[[Artifact]], [[Body Part]] [[Animate]]**	**[[Artifact]]**$_{AB}$ [[Area]]$_{A}$ **[[Body Part]]**$_{AB}$[[Location]]$_{AB}$ **[[Animate]]**[$_{AB}$ [Material]]$_{AB}$ [[Physical Object Part]]$_{AB}$ [[Surface]]$_{AB}$ [[Wind]]$_{AB}$ [[Psych]]$_{A}$ [[Building Part]]$_{B}$ [[Solid]]$_{B}$[[Inanimate]]$_{A}$ [[Route]]$_{B}$ [[Watercourse]]$_{B}$ [[Natural Landscape Feature]]$_{B}$ [[Stuff]$_{B}$] [[Sense Organ]]$_{B}$ [[Room]]$_{B}$
Pulire T-PAS2	Subj	**[[Human]]**	**[[Human]]**$_{AB}$
	Obj	**[[Body Part]] [[Garment]]**	**[[Body Part]]$_{AB}$ [[Garment]]**$_{AB}$ [[Animal]]$_{AB}$
Sbottonare T-PAS1	Subj	**[[Human]]**	**[[Human]]**$_{AB}$
	Obj	**[[Garment]]**	**[[Garment]]**$_{AB}$ [[Inanimate]]$_{A}$ [[Artifact]]$_{B}$ [[Cloth]]$_{B}$ [[Physical Object Part]]$_{AB}$
Sbottonare T-PAS2	Subj	**[[Human]]**	**[[Human]]**$_{AB}$
	Obj	**[[Garment]]**	**[[Garment]]**$_{AB}$ [[Artifact]]$_{B}$ [[Physical Object Part]]$_{AB}$
Sbottonare T-PAS3	Subj	**[[Garment]]**	**[[Garment]]**$_{AB}$
Sbottonare T-PAS4	Subj	**[[Human]]**	**[[Human]]**$_{AB}$ [[Institution]]$_{A}$ [[Human Group]]$_{A}$ [[Area]]$_{B}$
Vendere T-PAS1	Subj	**[[Human]] [[Business Enterprise]]**	**[[Human]]**$_{AB}$ [[Institution]]$_{AB}$ **[[Business Enterprise]]**$_{AB}$ [[Anything]]$_{B}$
	Obj	**[[Inanimate]] [[Animal]]**	**[[Inanimate]]$_{AB}$ [[Animal]]**$_{AB}$ [[Software]]$_{A}$ [[Asset]]$_{A}$ [[Wood]]$_{B}$[[Gas]]$_{B}$ [[Hair]]$_{A}$ [[Group]]$_{B}$ [[Abstract Entity]]$_{B}$ [[Priviledge]]$_{B}$ [[Numerical Value]]$_{B}$ [[Sense Organ]]$_{B}$ [[Information Source]]$_{B}$ [[Anything]]$_{B}$[[Energy]]$_{AB}$ [[Land]]$_{B}$ [[Area]]$_{B}$ [[Activity]]$_{AB}$ [[Body Part]]$_{AB}$ [[Physical Object Part]]$_{B}$
Vendere T-PAS2	Subj	**[[Human]], [[Business Enterprise]]**	**[[Human]]**$_{AB}$ [[Location]]$_{B}$ **[[Business Enterprise]]**$_{AB}$
Vendere T-PAS3	Subj	**[[Artwork]] [[Document]] [[Software]]**	**[[Artwork]]**$_{AB}$ **[[Document]]**$_{AB}$ **[[Software]]**$_{AB}$
Vendere T-PAS4	Subj	**[[Artifact]]**	**[[Artifact]]**$_{AB}$
Vendere T-PAS5	Subj	**[[Human]]**	**[[Human]]**$_{AB}$

is more specific than the expected one, namely [[Inanimate]]. In doing so, the annotator was able to retrieve a satisfactory description of the lexical item that in her opinion was somewhat underspecified.

Example 10
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"[..] l'occidente **vendeva** il gas a una nazione in lotta e le maschere antigas all'altra [..]"
Eng. '[..] Western countries used to **sell** gas to a nation at war and sell anti-gas masks to the other [..]"

Table 6 clearly shows that annotator B used a greater number of Semantic Types located at the bottom level of the ontology compared to annotator A, thus tagging the type associated with the entity denoted by the single lexical items instead of the generalization made by the T-PASs. For instance in Example 11 *legname* has been tagged as [[Wood]] by annotator B instead of [[Inanimate]].

Example 11
T-PAS#1 of the verb **vendere**
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"[..] si **vendeva** il legname [..]"
Eng. '[..] they used to **sell** woods '

This caused a sentence-specific annotation that employs Semantic Types that are leaf-nodes in the hierarchy, and as a result the generalization made by the T-PAS is lost. As future work, we plan to develop a methodology to normalize the Semantic Types to the appropriate level of abstraction in order to restore the generalization over lexical sets, which is a distinctive feature of the T-PAS resource.

We believe there are also linguistic reasons that intervene in the assignment of different Semantic Types to the same lexical element by the annotators. Annotators captured repeatedly the phenomenon known as regular polysemy (Apresjan 1974), systematic alternations of meaning that apply to classes of words instead of single words (Jezek 2016). [11] One annotator tagged an argument according to the basic meaning of the polysemy pattern whereas the other tagged the same argument according to the other one. For instance both Example 12 and 13 record the systematic alternation between [[Artifact]] and [[Route]] that is typical of lexical items like *strade* (Eng. 'streets') and *sentiero* (Eng. 'path'), which has been tagged differently by annotators but accordingly to the regular polysemy alternation.

Example 12
T-PAS#1 of the verb *pulire*
[[Human]-subj] **pulire** [[Body Part | Artifact | Animate = Food]-obj]
"La causa per cui le imprese non hanno prontamente **pulito** le strade [..]"
Eng. 'The cause for which the enterprises did not **clean** the streets [..]'

Example 13
T-PAS#1 of the verb *pulire*
[[Human]-subj] **pulire** [[Body Part | Artifact | Animate = Food]-obj]
"Luis, in testa alla fila, **puliva** il sentiero col macete [..]"

11 Apresjan (1974) identified several patterns of regular polysemy of nouns and other parts of speech.

Eng. 'Luis, ahead of the group, was **clearing** the path with the machete'

Another systematic alternation recorded by annotators is the one shown in Example 14, where *terre* (Eng. 'lands') has been tagged as [[Inanimate]] by one annotator and [[Land]] by the other. [[Land]] is a Semantic Type belonging to the [[Location]] branch of the hierarchy, which is very different from the [[Inanimate]] one.

Example 14
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"Per sollevare il bilancio dello stato decise di **vendere** le terre demaniali [..]"
Eng. 'In order to rise the state balance he decided to **sell** his lands [..]'

In addition, annotators marked tokens like *rulli* (Eng. 'gears') and *lente* (Eng. 'lents') in Example 15 and 16 respectively, both as [[Physical Object Part]] and [[Artifact]], according to the well-known *part* for *whole* alternation.

Example 15
T-PAS#1 of the verb *pulire*
[[Human]-subj] **pulire** [[Body Part ⁞ Artifact ⁞ Animate = Food]-obj]
"[..] per **pulire** in modo più accurato i rulli dell'impastatrice dai residui di impasto [..]"
Eng. ' In order to clean accurately the gears of the dough maker from the residue of the dough [..] '

Example 16
T-PAS#1 of the verb *pulire*
[[Human]-subj] **pulire** [[Body Part ⁞ Artifact ⁞ Animate = Food]-obj]
"Giles **pulisce** una lente dei suoi occhiali."
Eng. 'Eng.'Giles **cleans** a lens of his glasses.'

Finally, there are differences in ST assignment due to inherent polysemy. 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. Examples are provided by Example 10 and Example 17 where *Occidente* (Eng. 'Western Countries') and *Stato* (Eng. 'State') have been tagged by one annotator as [[Business Enterprise]] and by the other as [[Area]]. Note that in the specific context [[Area]] is not the appropriate annotation.

Example 17
T-PAS#1 of the verb *vendere*
[[Human | Business Enterprise]-subj] **vendere** [[Inanimate | Animal]-obj]
"Lo Stato [..] del resto non vende le sigarette?"
Eng. 'The State [..] after all doesn't it **sell** sigarettes?'

6. Related work

The effort of semantic type tagging presented in this paper relates to the attempts in the computational linguistics literature to assess selectional preferences and to estimate thematic fit for verbs (semantic plausibility of words to fill verb argument roles) within

the distributional semantic paradigm, in particular, to the work of (Erk 2007) and more recently (Sayeed and Demberg 2014). While the majority of the works devoted to thematic fit estimation addresses the problem of assessing the semantic constraints a verb poses on its arguments in terms of cognitive expectation and psycholinguistic processing, in T-PAS we collect human judgments on semantic type selection based on corpus analysis and annotation, with the goal of providing a gold standard for linguistic analyses and natural language processing tasks.

7. Conclusions

We performed a pilot experiment to tag the arguments of verbs, as recorded in the T-PAS resource, with their associated Semantic Type. The main goal of the annotation exercise was to design an annotation task that could be used for the expansion of the annotation of the corpus of the T-PAS resource. In order to achieve this goal we performed an annotation exercise and ran an Inter-Annotator Agreement test on part of the annotated data. By analyzing the cases of inter annotator disagreement, we were able to identify phenomena that on the one hand affect the Inter-Annotator Agreement score, while on the other hand represent instances of aspects to improve in our annotation design. Overall, we obtained good results in the annotation. Future works include spelling out the rules for polysemous words tagging more clearly in the guidelines, employing the developed annotation scheme for the systematic extension of the annotation and also implementing a semantic parser for Italian using the annotated data as a starting point.

Acknowledgments

We acknowledge the reviewers for their helpful comments. Their observations were useful to improve the article.

References

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

Baroni, Marco 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 (EACL '06)*, pages 87–90, Trento, Italy, April 3-7. Association for Computational Linguistics.

Bartalesi Lenzi, Valentina, 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, Istanbul, Turkey, May 21-27.

Cinková, Silvie, 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, Istanbul, Turkey, May 21-27.

Erk, Katrin. 2007. A simple, similarity-based model for selectional preferences. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 216–223, Prague, Czech Republic, June 23-30.

Hanks, Patrick. 2004. Corpus pattern analysis. In *Proceedings of the Eleventh EURALEX International Congress*, pages 87–97, Lorient, France, July 6-10.

Hanks, Patrick and Elisabetta Jezek. 2008. Shimmering lexical sets. In *Proceedings of the XIII EURALEX International Congress*, pages 391–402, Barcelona, Spain, July 15-19.

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

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

Jezek, Elisabetta, Anna Feltracco, Lorenzo Gatti, Simone Magnolini, and Bernardo Magnini. 2016. Mapping semantic types onto wordnet synsets. In *Twelfth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (Isa-12)*, pages 11–15, Portorose, Slovenia, May 23-28.

Jezek, Elisabetta and Francesca Frontini. 2010. From pattern dictionary to patternbank. *A Way with Words: Recent Advances in Lexical Theory and Analysis, Kampala, Menha Publishers*, pages 215–239.
Jezek, Elisabetta and Patrick Hanks. 2010. What lexical sets tell us about conceptual categories. *Lexis*, 4(7):22.
Jezek, Elisabetta, 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)*, Reykjavik, Iceland, May 26-31.
Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
McCarthy, Diana and John Carroll. 2003. Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, 29(4):639–654.
Ponti, Edoardo Maria, Elisabetta Jezek, and Bernardo Magnini. 2016. Grounding the lexical sets of causative-inchoative verbs with word embedding. In *Proceedings of the Third Italian Conference on Computational Linguistic (CLiC-it 2016)*, Napoli, Italy, December 5-6.
Ponti, Edoardo Maria, Elisabetta Jezek, and Bernardo Magnini. 2017. Distributed representations of lexical sets and prototypes in causal alternation verbs. *Italian Journal of Computational Linguistics*, 3(1):25–42.
Pustejovsky, James. 1995. *The generative lexicon*. The MIT Press.
Resnik, Philip. 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, Washington, D.C., USA, April 4-5.
Rijsbergen, CJ van. 1979. Information retrieval. 1979.
Rosch, Eleanor H. 1973. Natural categories. *Cognitive psychology*, 4(3):328–350.
Sayeed, Asad and Vera Demberg. 2014. Combining unsupervised syntactic and semantic models of thematic fit. In *Proceedings of the first Italian Conference on Computational Linguistics (CLiC-it 2014)*, Pisa, Italy, December 9-10.