Enriching 'Senso Comune' with Semantic Role Sets

Elisabetta Jezek, Laure Vieu, Fabio M Zanzotto, Guido Vetere, Alessandro Oltramari, Aldo Gangemi, Rossella Varvara

Enriching 'Senso Comune' with Semantic Role Sets

E. Jezek[1], L. Vieu[2], F.M. Zanzotto[3], G. Vetere[4], A. Oltramari[5], A. Gangemi[6], and R. Varvara[7]

(1) Università di Pavia, (2) IRIT-CNRS - Université Toulouse III,
(3) Università di Roma "Tor Vergata", (4) IBM Center for Advanced Studies,
(5) Carnegie Mellon University, (6) ISCT - CNR, (7) Università di Trento
jezek@unipv.it, vieu@irit.fr, fabio.massimo.zanzotto@uniroma2.it, gvetere@it.ibm.com,
aoltrama@andrew.cmu.edu, aldo.gangemi@cnr.it, rossella.varvara@unitn.it

Abstract

The paper describes the design and the results of a manual annotation methodology devoted to enrich the *Senso Comune* resource with semantic role sets for predicates. The main issues encountered in applying the annotation criteria to a corpus of Italian language are discussed together with the choice of anchoring the semantic annotation layer to the underlying dependency syntactic structure. We describe the two experiments we carried to verify the reliability of the annotation methodology and to release the annotation scheme. Finally, we discuss the results of the linguistic analysis of the annotated data and report about ongoing work.

1. Introduction

Large-scale linguistic resources that provide relational information about predicates and their arguments are indispensable tools for a wide range of NLP applications, where the participants of a certain event expressed by a predicate need to be detected. In particular, hand-annotated corpora combining semantic and syntactic information constitute the backbone for the development of probabilistic models that automatically identify the semantic relationships conveyed by sentential constituents in text, as in the case of Semantic Role Labeling (Gildea and Jurafsky, 2002). In addition, annotated corpora enable the quantitative and qualitative study of various linguistic phenomena at the syntax-semantics interface and the development of data-driven models for lexical semantics.

The LIRICS (Linguistic Infrastructure for Interoperable ResourCes and Systems) project has recently evaluated several approaches for semantic role annotation (PropBank, VerbNet, FrameNet, among others) and proposed an ISO (International Organization for Standardization) ratified standard for semantic role representation that enables the exchange and reuse of (multilingual) language resources. The standard comprises 29 'high level' (coarse-grained) roles identified using an entailment-based methodology (Petukhova and Bunt, 2008; Gotsoulia 2011). This set has been mapped (*inter alia*) onto VerbNet roles and organized hierarchically (Bonial et al. 2011 a, b). Similar lexicons/annotation efforts include the German SALSA project (Burchardt et al. 2006), the Czech dependency treebank and its PDT-Vallex valency lexicon.

In this paper we present the design and the results of a manual annotation methodology based on the ISO-semantic roles, aiming at enriching the *Senso Comune* knowledge base of the Italian language (henceforth SC) with semantic role sets for predicates, to be used for linguistic research and NLP applications. In SC semantic roles sets are not assigned to predicates axiomatically but they are induced by the annotation of the usage examples associated with the *sensi fondamentali* (word meanings which are predominant in terms of use among the most frequent 2000 words in the language, cf. De Mauro, 1999) of the verb lemmas. The methodology encompasses annotation of the role played by participants in the event described by the predicate (intentional agent, affected entity, created entity and so on) as well as annotation of their inherent semantic properties, expressed in the form of ontological categories (person, substance, artifact, and so forth).

In the rest of the paper, we first present an overview of the SC resource, then introduce the annotation scheme and the experimental setting in which the scheme was finalized. Finally, we discuss the results of the annotations in terms of inter-annotator agreements and linguistic generalizations that can be drawn form the analysis of the data. We conclude by observing how interoperability of lexical data can also be supported formally (in the spirit of SC) in a linked data perspective.

2. Resource overview

The SC model features the main structures of standard lexicography (we refer to Vetere et al. 2012 for a general overview). These consist in lexical entries (lemmas) with their linguistic characterization and their senses. Each sense is comprised of a definition (glossa), a number of usage marks, specific grammatical constraints, usage instances, and lexicographic relations. In addition, SC provides substantive senses with ontological annotations, whose labels are taken from a foundational ontology inspired to DOLCE (Gangemi et al. 2002). The idea at the basis of ontological annotations is that linguistic senses (also referred to as *linguistic concepts*) are *tangential to reality*: they are abstract *social* entities whose relationship with extra-linguistic realities is established in the context of human activities. This idea, which comes from semiotics, calls for a formal distinction between two kinds of intensional entities: linguistic concepts (i.e. senses) and ontological categories. In fact, the ontological classification of linguistic concepts is not intended as a direct extensional interpretation over some domain of *real entities*. Instead, we resort on a notion of *ontological commitment*: a word can be used in a certain sense to refer (even vaguely, evocatively, notionally or metaphorically) to entities of some hypothetical kind.

Also, we adopt the distinction between *type* and *token* which comes form classic semiotics (Peirce); the former being abstract sorts, the latter their situated concrete instances. For instance, the Gertrude Stein's verse *a rose is a rose is a rose* counts three *rose* word tokens which instantiate FLOWER-ROSE, i.e. the (single) specific sense of *rose* occurring in the sentence, which, in turn, *commits* to the existence of objects which fall under the NATURAL-OBJECT ontological category. Note that *commits* is not to be read as *logical implication*; on the contrary, senses and ontological categories are logically disjoint, so that lexical relationships (e.g. synonymy) do not imply, nor conflict with, ontological axioms (e.g. equivalence).

3. Annotation scheme and methodology

On approaching the task of providing SC with verbal frames, we decided to start from tokens instead of types. Rather than speculating about predicate structures associated with verbal senses, we focused on annotating usage instances, as registered in the dictionary. The compilation of type-level verbal frames *à la* VerbNet is therefore deferred to a later process of generalization.
To encode the annotation of verbal predicate structures, we opted for a model based on dependencies between shallow syntactic structures, inspired to eXtended Dependency Graphs (XDG) (Basili and Zanzotto, 2002). Basically, the scheme foresees:

- the identification of flat constituents (chunks)
- the identification of the verbal chunk which conveys the exemplified sense
- the annotation of phrases which hold a thematic relation with the verb.

Argumental phrases are annotated according to the following characterization:

- each argumental chunk is given
  - a syntactic role (e.g. SUBJECT)
  - a constituent type (e.g. NP)
  - a semantic role (e.g. AGENT)
  - an ontological category (e.g. HUMAN)
- tokens of the argumental chunk are
  - (automatically) assigned a POS tag and a lemma (lemmatisation)
  - (optionally, and manually) assigned a sense (disambiguation)

Both lemmatisation and disambiguation are based on the SC dictionary. The information structure described above is encoded in a specific annotation data model (Fig. 1). This model is specified in OWL, as part of the ontology underlying the SC knowledge base [1]. Also, we provide a Java implementation which is made persistent and accessible

[1] http://www.sensocomune.org/ontologies/

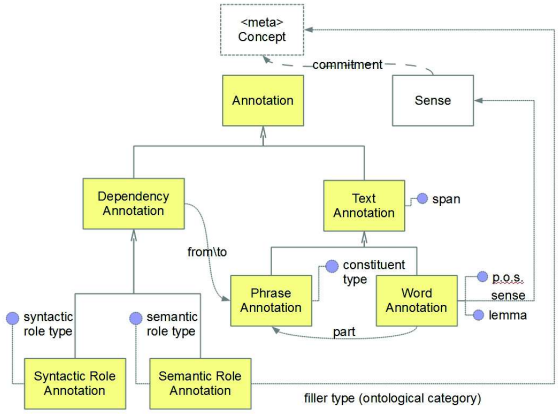

Figure 1: The Annotation Model

on relational databases through an object-relational mapping. Thus, actual annotation data are integrated in the general SC database, which allows issuing conjunctive queries where lemmas, senses, grammatical features and argument structures can be joined to extract relevant patterns.
The induction of type-level verbal frames from usage annotation data will require a process of generalization whose study is included in our future plans. To represent typical verbal frames, we plan to adopt a model in which semantics and syntactics are structurally separated, and yet logically connected. This model aims at preserving the generality of semantic structures as distinct from their syntactic realizations. Our intuition is that, by decoupling semantic and syntactic frames, one could achieve a powerful and concise representation of linguistic data, to better handle and investigate their interplay. For instance, action frames including participants and objects may be rendered in either passive or active forms; still, retrieving the lexical concepts involved in certain actions can abstract from the syntactic unfolding of verbal arguments.
In the following sections we describe the component and tags of the scheme in more detail.

3.1. Constituents and Dependency relations

We choose a light annotation scheme for syntactic dependency relations. Focusing the attentions to the verb dependency relations, we defined three types of relations: Subject (S), Object (O), and other Complement (C). We avoided the distinction, at the syntactic level, between Complement and Adjunct. This distinction is out of the scope of the syntactic phase as it is a target of the overall process of frame annotation.
As the model is inspired to the extended dependency graphs XDG) (Basili and Zanzotto, 2002), the syntactic dependency relations link constituents. We focus on the constituents that may play a role as verb arguments: Nominal Phrases (Sintagma Nominale, SN), Pronoun Phrases (Sintagma Pronominale, Spron), Prepositional Phrases (Sintagma Preposizionale, SPrep), Adverbial Phrases (Sintagma Avverbiale, SAvv), Adjectival Phrases (Sintagma Aggettivale, SAgg), and SubSentence (Sottofrase, SFr). This latter is little tricky as it is defined as a subsentence headed by a verb that is not the target verb. An example for

SC role	LIRICS role
Agente (AG)	Agent, Partner
Causa (CAUSE)	Cause, Reason
Strumento (INSTR)	Instrument, Means
Paziente (PT)	Patient
Tema (TH)	Theme, Pivot
Goal (GOAL)	Goal
Beneficiario (BEN)	Beneficiary
Origine (SOURCE)	Source
Luogo (LOC)	Location, Setting
LuogoFinale (ENDLOC)	EndLocation
LuogoIniziale (INITLOC)	InitialLocation
Percorso (PATH)	Path
Distanza (DIST)	Distance
Tempo (TIME)	Time
TempoFinale (ENDTIME)	EndTime
TempoIniziale (INITTIME)	InitialTime
Durata (DUR)	Duration
Risultato (RESULT)	Result
Quantità (AMOUNT)	Amount
Maniera (MANNER)	Manner, Medium
Esperiente (EXP)	Pivot, Patient
Scopo (PURPOSE)	Purpose
Frequenza (FREQ)	Frequency
Attributo (ATTR)	Attribute

Table 1: Semantic roles set

the two levels of annotations is the following:

Const. (SN Luca) ha dedicato (SN il libro) (SPrep alla madre)
Dep. (S Luca) ha dedicato (O il libro) (C alla madre)
(Luca dedicated a book to his mother)

where *Luca* and *il libro* (the book) are nominal phrases (SN) and *alla madre* (to his mother) a prepositional phrase (SPred). The three phrases play, respectively, the syntactic role of subject (S), object (O), and other complement (C).

3.2. Semantic Role list

The list of SC roles comprises 24 coarse-grained (high-level) semantic roles based on LIRICS (Petukhova and Bunt 2008) and the on-going attempt to create a unified standard set for the International Standard Initiative with the goal of facilitating mappings between semantic resource of different granularity, including VerbNet (Bonial et al. 2011 a, b). In designing the set, we conflated some LIRICS roles such as Agent and Partner (Co-Agent in VerbNet), and used some classical semantic roles like Experiencer rather than LIRICS's ambiguous Pivot. The final set of categories is given in Table 1, together with the mappings with the ISO roles of LIRICS. Each roles is defined by a gloss and a set of examples, in the LIRICS style.

3.3. Role Taxonomy

To facilitate the understanding of the scheme adopted, in addition to the glosses and the examples, semantic roles are structured into the taxonomic hierarchy of Fig. 2, in a similar way to what is done in (Bonial et al. 2011b) for LIRICS and VerbNet unified roles.
A main difference is that we have added intermediate nodes that do not count as role labels, but, with further glosses, help the annotator in understanding the main discriminating elements between roles. This enabled implementing an ontological distinction between roles that identify event participants proper, and roles that identify elements of the context of the event. As a result, some distinctions that might be difficult to grasp at first, such as Luogo Iniziale (Initial Location) vs. Origine (Source), are made clearer: in this example the first is part of the spatial context of the event, while the second is a proper and non-spatial participant to the event.

3.4. Ontological categories and TMEO methodology

In the context of *Senso Comune* we developed a tutoring system to support collaborative ontology population. As the acronym may suggest to philosophers, TMEO (Tutoring Methodology for the Enrichment of Ontologies) recalls Plato's dialectic methodology of discovering knowledge through reasoning in dialogues (Reale 1990): in this regard, by distilling the key ontological properties of SC into germane questions targeted at users, TMEO plays the role of a 'digital Socrates' in a basic interaction system. For instance, consider the scenario in which a given user is asked to classify the term *shoe*, in the sense of "footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material". TMEO system's interface will submit a series of intuitive conceptual questions to the users in order to disambiguate the intended meaning of the term. The following sequence represents a simplified scenario based on this example:

- TMEO: Can you touch, see, smell, taste, feel **a shoe**? User: Yes
- TMEO: Would you say that "a **shoe** can happen or occur? User: No
- TMEO: In general, does it make sense to use the word **shoe** as answer to the question "when"? User: No
- TMEO: does **shoe** indicate a location? User: No
- TMEO: Can **shoe**s act by intention? User: No
- TMEO: Would you say that **shoe**s are built by someone? User: Yes
- TMEO: **shoe** in the sense of 'footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material' has been classified as ARTIFACT.

As the above-mentioned scenario suggests, TMEO methodology may therefore be adopted not only in the unilateral classification of a given term ('shoe') but also in making related lexical items explicit. This kind of relatedness between terms actually unwraps the inter-categorial relation(s) holding between the corresponding ontological categories (since a detailed presentation of TMEO is out of scope in the current paper, we remand the reader to a more comprehensive publication (Oltramari et al. 2012).

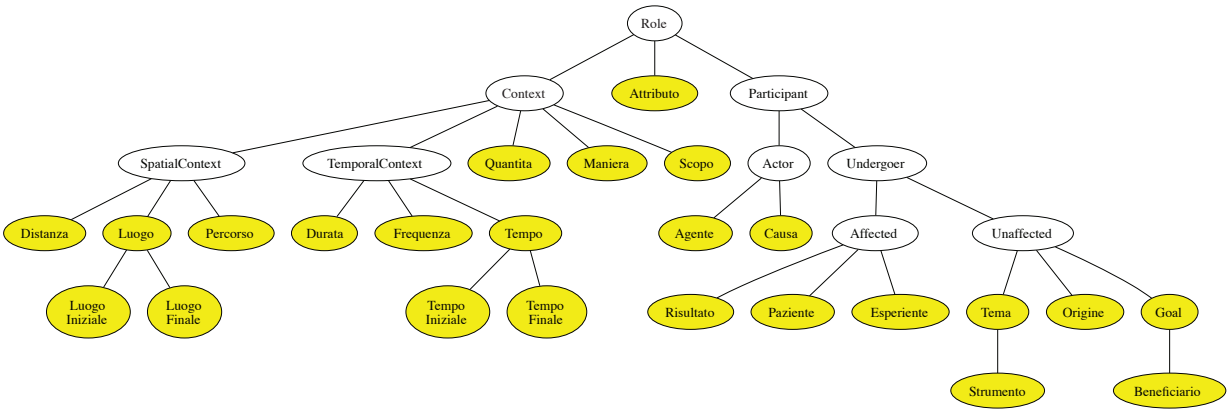

Figure 2: The semantic role taxonomy.

TMEO has been implemented as a finite state machine (FSM): in general, the elaboration process of a FSM begins from one of the states (called a 'start state'), goes through transitions depending on input to different states and must end in any of those available (only the subset of so-called 'accept states' mark a successful flow of operation). In the architectural framework of TMEO, the 'start state' is equivalent to the top-most category ENTITY, the 'transitional states' correspond to disjunctions within ontological categories and 'accept states' are played by the most specific categories of the model, i.e. 'leaves' of the relative taxonomical structure. In this context, queries represent the conceptual means to transition: this means that, when the user answers to questions like the ones presented in the above-mentioned example, the FSM shifts from one state to another according to answers driven by boolean logic[2]). If no more questions are posited to the user, this implies that the system has reached one of the available final 'accept state', corresponding to the level where ontological categories don't have further specializations. TMEO human language interface is very intuitive and comes in the form of a map where *yes/no* options are presented together with the step-by-step questions: figure 3 shows the 'shoe' example in the Italian translation 'scarpa'. In future work we aim at extending the coverage of TMEO's model and improving the scalability of the system towards genuine crowd-based platforms.

The ontological categories underlying the TMEO methodology form a taxonomy as in Fig. 4.

The annotation of ontological categories performed in the context of the work reported here differs from the annotations already present in the SC resource and described in earlier work. Here, instead of a lexical entry with its gloss, annotators were presented a text span in the context of a usage instance. In addition, they were suggested to annotate this text span with multiple categories if this was deemed more adequate than a single one. Such a possibility was introduced to acknowledge the inadequacy of a unique categorization when several

[2]Uncertainty will be included only in future releases of the TMEO system.

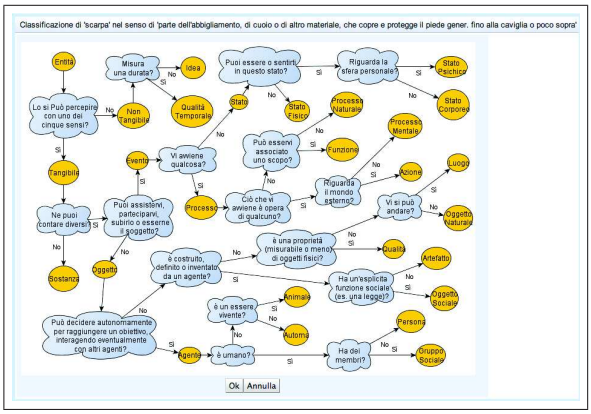

Figure 3: Senso Comune's interface for TMEO

interpretations co-exist due to systematic polysemy (e.g. "book" often refers simultaneously to an artifact and to an information object). Finally, the annotators were pushed to distinguish between singular and collective use of such categories. As a result, a text span like "Un ufficio" in the example "Un ufficio che funziona" ('An office that works well') can possibly be annotated POSTO+PERSONA COLLETTIVO+ORGANIZZAZIONE (Place+PersonCollective+Organization).

4. Annotation reliability

We verified the reliability of the annotation scheme by comparing annotations carried out by multiple annotators independently. In the following sections we describe the two pilot experiments we carried out, during which the same portion of the corpus was annotated by several participants.

4.1. Annotation experiment

We evaluated the annotation procedure in two experimental settings involving multiple annotators and estimated their agreement on the task. We selected 22 target verbs and performed multiple annotation on a set of 66 non disambiguated examples (3 for each target verb). The annotation

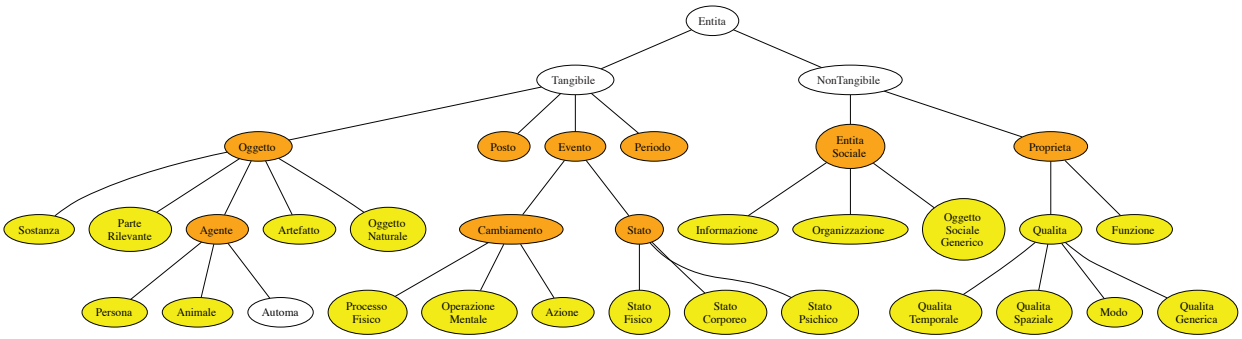

Figure 4: The ontological category taxonomy.

task was split in two subtasks. We first performed syntactic and semantic role annotation; then, we supplied the annotators with the data annotated with the sole syntactic layer, and asked them to annotate the ontological category of the argument fillers. Verbs were selected according to variability in semantic selection (for both roles and ontological categories) and syntactic realization.

4.2. Span detection

Detection the span of the verb arguments is one of the most important activity when annotating. The span of the verb argument define the sentence chunk that has to be syntactically and semantically annotated. Each annotator has to work on the same span in order to make annotations comparable. Even if the annotators decide for the same syntactic and semantic label for a nearly similar chunk of sentence, annotations cannot be compared. Thus, for comparing the annotations we assessed a gold standard, that is the most voted span for each argument.

5. Results

5.1. Interannotator agreement

The two annotation experiments were done by 9 annotators each. Among those annotators, we removed a few outliers, 1 in the first experiment and 2 in the second, for obvious misunderstanding of the task, resulting in 8 and 7 annotators respectively. We chose to use average pairwise Cohen's kappa as a measure of inter-annotator agreement, data being particularly skewed (Artstein and Poesio 2008).

For the first experiment, the inter-annotator agreement among the 8 annotators is 0.86 for the subtask on syntactic dependency relations (4 labels: 3 relations + no annotation) and 0.66 for the subtask on semantic roles (25 labels: 24 roles + no annotation). Such values are usually considered respectively as very good and fair, the latter especially so since semantic tasks are notoriously difficult.

Subgroups of annotator apparently achieved a deepest expertise, with pair agreement respectively reaching maximums of 0.91 and 0.88 on each sub-task.

In the second experiment, since we gave annotators the possibility to annotate multiple categories, there were in total 60 different labels (including no annotation). The raw agreement among the 7 annotators is quite low at 0.41. Taking into account partial agreement in the relatively few cases in which annotators used multiple categories (27 occurrences) and/or used the collective tag (36 occurrences), the agreement slightly rises to 0.46, with a pairwise maximum of 0.57. However, taking advantage of the hierarchical organization of the categories into a taxonomy, meaningful aggregation of categories can be proposed. For instance, one can reduce the 30 base-category labels in Fig. 4 actually used (only the coloured nodes have been used in the experiment), a rather large figure, into 9 labels corresponding to the orange-coloured ones on this figure. This forms a more shallow ontology, but still a meaningful discriminating one, and yields 17 different labels (with multiple categories and collectives). With such a reduction of the labels, the overall agreement clearly increases at a reasonable 0.60, with a pairwise peak at 0.79. Further analysis of the data may show where exactly annotators tend to diverge, enabling focusing on specific merges only and keeping a more fine-grained taxonomy.

6. Linguistic analysis of annotations

Besides confirming well-known difficulties in semantic role annotation, such as confusion between PT and TH due to uncertainties in the interpretation of the notions of "modification", the specificity of the annotation scheme allows us to make interesting observations regarding the role played by the semantic context, particularly the ontological category associated with the argument filler, in semantic roles annotation. This can be illustrated by focusing on the annotation of the semantic role of the subject for the 24 cases in our corpus in which there is complete agreement about the inanimate nature of referent of the filler. The first observation is that in these cases there is much more confusion between roles than average (average of kappa = 0,51). In our view this is related to the following aspects (as a reference theoretical framework cf. Pustejovsky 1995):

- there is metonymy between verb and argument in the context
- the noun is inherently polysemous
- the verb exhibits a shift in meaning
- the annotator confuses the inherent properties of the argument filler with its role.

Consider for example the case of disagreement between AG and TH (the most frequent in this set of data), that can be found in examples such as "il treno corre nella pianura a 100 all'ora" ('the train runs in the plains at 100 Km/h' 3AG / 5TH). In these cases, the annotator is confused by the fact that the verb in its basic meaning reports an intentional eventuality, whereas the filler in the instance is inanimate. It appears that two solutions are taken in annotation: either the filler is somewhat interpreted metonymically and assigned the AG role, or the verb is interpreted as carrying a meaning which is not the basic agentive meaning, and the subject is tagged TH.

The additional case of "Un ufficio che funziona" ('An office that works well' 5 AG / 3 TH) appears to be more complex, due to the inherent polysemy in the noun. In fact, in this case, we register high disagreement not only at the level of roles but also at the level of ontological categories, where *ufficio* is annotated as POSTO ('place', 2/7 annotators), ORGANIZZAZIONE ('organization', 2/7), PERSONA COLLETTIVO ('person collective', 2/7), POSTO+PERSONA COLLETTIVO ('place+people', 1/7).

In this case, one can argue that two phenomena are at play simultaneously, which confuse the annotators: the verb disambiguates the polysemous noun in context but at the same time its meaning is redefined by it (from 'to work properly' to 'to perform a task well').

Among our 24 cases, other significant cases of disagreement can be found with nouns denoting instruments. Consider the examples "la penna scrive nero" 'the pen writes black' and "forbici che tagliano bene" 'scissors that cut well', that have been annotated as INSTR by 3/8 and 4/8 respectively (*pen* was further tagged as TH by 5/8, while *scissors* as TH by 3/8 and AG by 1/8). These subjects (called *Instrument subjects* in literature, see e.g. Alexiadou et Schäfer 2006) refer to entities frequently used as facilitating instruments in everyday life (as expressed in sentences like "I wrote the letter with a fountain pen", "I used the scissors to open the package"), but in the examples above they are not presented as instruments, but rather as the entity about which the verb predicates something (that is, they have the characteristic of writing and cutting). Nobody uses them to perform an action; hence, they are THs because they are the participants in the condition described by the verb and are not modified by the event. We argue that in these cases annotators who tag them INSTR confuse the ontological type of the entity denoted by the filler with the semantic role the participant plays in the event.

7. Interoperability of Semantic Roles on the Semantic Web

SC has been formally represented in OWL, and this offers an opportunity to make it interoperable at both synset level (through an ongoing alignment to the Italian version of MultiWordNet, which will be part of the Lexical Linked Data Cloud), and at semantic role level, by aligning it to the VerbNet and FrameNet RDF datasets.

Recently, the problem of interoperability between different linguistic ontologies (schemas for representing linguistic data) has entered the Semantic Web and Linked Open Data radar, since there are mutual advantages in creating linguistic data expressed in RDF (the basic language for the Semantic Web): the Web as an integration platform for heterogeneous linguistic data, as well as easier support for lexicalizing ontologies.

In that context, several initiatives are boosting the adoption of good practices for sharing linguistic data, and make them interoperable at a formal level. NLP Interchange Format (NIF) is an RDF/OWL-based format that allows to combine and chain several NLP tools in a flexible, light-weight way. The Linguistic Linked Open Data initiative is linking many linguistic datasets, but it is still missing a tight integration of lexical resources including semantic roles. FrameNet and VerbNet have been ported to RDF and OWL (cf. Nuzzolese et al. 2011 for FrameNet-OWL), including the mapping between FrameNet frames and VerbNet predicates, but this is not yet extended to the respective role structures. The OntoLex W3C Community Group is going to publish a proposal for a standard to describe lexical resources jointly with ontologies and linked datasets (where the basic innovation is to allow for a sense layer distinguished from lexical expressions and ontological entities, which enables intensional semantics of lexical resources to be used in the mostly extensional formal semantics assumed in the Semantic Web).

The potential of the Semantic Web for semantic role labeling (and vice versa) is exemplified by the FRED architecture (Presutti et al. 2012), where VerbNet roles are used to automatically annotate RDF graphs that are extracted from text by means of multiple NLP algorithms (semantic role labeling, frame detection, relation extraction, sense disambiguation, named entity recognition).

FRED allows to link those graphs to linked data resources; it aligns named entities to linked data resources, as well as named concepts (typically derived from disambiguated terms) to WordNet or DBpedia resources. Since RDF resources are usually typed, FRED graphs can be used for investigating the actual coverage of VerbNet roles, with their associated types (à la selectional restrictions). In fact, FRED complements partial coverage of VerbNet with other roles, e.g. directly expressed by prepositions, which can be further investigated.

8. Conclusions

In this paper, we described the design of a manual annotation methodology devoted to enrich the SC resource with semantic role sets for predicates. We discussed the results of the two experiments performed to verify the reliability of the annotation methodology, in terms of inter-annotator agreement and linguistic generalizations that can be drawn form the analysis of the data. For the future, we plan to perform automatic chunking of the data to be annotated and check it manually before annotation; to annotate the ontological category of the argument fillers out of context; to develop a methodology for extraction of semantic roles sets for predicates from the annotated data; to link SC semantic roles sets to other lexical resources for Italian such as T-PAS structures (Jezek et al. 2014).

Acknowledgments

We thank three anonymous reviewers for their useful comments.

References

A. Alexiadou and F. Schäfer. 2006. Instrument Subjects Are Agents or Causers. In D. Baumer, D. Montero, and M. Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 40-48. Somerville, MA: Cascadilla Proceedings Project.

A. Burchardt , E. Katrin , A. Frank , A. Kowalski , S.Padó, M. Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. In Proceedings of LREC 2006.

R. Artstein and M. Poesio. 2008. Inter-coder agreement for computational linguistics. In *Computational Linguistics*. 34, 4, 555-596.

R. Basili and F.M. Zanzotto. 2002. Parsing engineering and empirical robustness. *Natural Language Engineering*, 8/2-3.

C. Bonial, S.W. Brown, W. Corvey, V. Petukhova, M. Palmer, H. Bunt. 2011a. An Exploratory Comparison of Thematic Roles in VerbNet and LIRICS. In *Proceedings of the Sixth Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*.

C. Bonial, W. Corvey, M. Palmer, V. Petukhova, H. Bunt. 2011b. A hierarchical unification of LIRICS and VerbNet semantic roles. In *Proceedings of the 2011 IEEE Fifth International Conference on Semantic Computing*, IEEE Computer Society Washington, DC, USA, 483-489.

T. De Mauro. 1999. Introduzione. In De Mauro T. (Ed. in Chief), *Grande Dizionario Italiano dell'Uso (GRADIT)*, 6 voll. + CD-rom, Torino, UTET, vol. I, VI-XLII.

D. Dowty. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67.3, 547-619.

A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider. 2002. Sweetening Ontologies with DOLCE. *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW 02)*.

D. Gildea and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245-288.

V. Gotsoulia. 2011. An Abstract Schema for Representing Semantic Roles and Modelling the Syntax-Semantic Interface. *Proceedings of the Ninth International Conference on Computational Semantics (IWCS '11)*, Association for Computational Linguistics Stroudsburg, PA, USA, 115-124.

E. Jezek, B. Magnini, A. Feltracco, A. Bianchini and O. Popescu. 2014. T-PAS: A resource of corpus-derived Types Predicate-Argument Structures for linguistic analysis and semantic processing. in *Proceedings of LREC 2014* (to appear).

A.G. Nuzzolese, A. Gangemi, V. Presutti. 2011. Gathering Lexical Linked Data and Knowledge Patterns from FrameNet. In O. Corcho, M. Musen (eds.) *Proceedings of K-CAP 2011 The Sixth International Conference on Knowledge Capture*, ACM.

A. Oltramari, A. Mehler, K. Kühnberger, H. Lobin, H. Lüngen, A. Storrer, A. Witt. 2012. An Introduction to Hybrid Semantics: The Role of Cognition in Semantic Resources. In *Modeling, Learning, and Processing of Text Technological Data Structures*, Berlin / Heidelberg, Springer, vol. 370, 97-109.

V. Petukhova, H. Bunt. 2008. LIRICS semantic role annotation: Design and evaluation of a set of data categories. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 28–30.

V. Presutti, F. Draicchio and A Gangemi. 2012. Knowledge Extraction based on Discourse Representation Theory and Linguistic Frames. In A. ten Teije and J. Vlker (eds.) *Proceedings of the Conference on Knowledge Engineering and Knowledge Management* (EKAW2012), LNCS, Springer.

J. Pustejovsky. 1995. *The Generative Lexicon*. Cambridge MA: MIT Press.

G. Reale. 1990. *A History of Ancient Philosophy: Plato and Aristotle*. SUNY Press.

G. Vetere, A. Oltramari, I. Chiari, E. Jezek, L. Vieu, F.M. Zanzotto. 2012. 'Senso Comune': An Open Knowledge Base for Italian. In *Revue TAL (Traitement Automatique des Langues)*, Journal Special Issue on *Free Language Resources* 52.3, 217-43.