Senso Comune, an Open Knowledge Base for Italian

Guido Vetere[*]; **Alessandro Oltramari**[**]; **Isabella Chiari**[***]; **Elisabetta Jezek**[****]; **Laure Vieu**[*****]; **Fabio Massimo Zanzotto**[******]

[*] *IBM Center for Advanced Studies of Rome, Via Sciangai 54, 00144 Roma, Italia*
[**] *Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue, 15213 Pittsburgh (PA), USA.*
[***] *Università La Sapienza di Roma, Dipartimento di Scienze Documentarie, linguistico-filologiche e geografiche, pl. Aldo Moro 4, 00185 Roma, Italia.*
[****] *Università di Pavia, Dipartimento di Linguistica Teorica e Applicata, Strada Nuova 65, 27100 Pavia, Italia.*
[*****] *IRIT-CNRS, Université Paul-Sabatier, 118 route de Narbonne, 31062 Toulouse, France & LOA-ISTC-CNR, Trento, Italia.*
[******] *Università di Roma "Tor vergata", Dipartimento di Ingegneria dell'Impresa, Viale del Politecnico, 1, 00133 Roma, Italia.*

ABSTRACT. Senso Comune is an open-knowledge base for the Italian language, available through a Web-based collaborative platform, whose construction is in progress. The resource integrates dictionary data coming from both users and legacy resources with an ontological backbone, which provides foundations for a formal characterization of lexical semantic structures (frames). A nucleus of basic Italian lemmas, which have been semantically analyzed and classified, is available for both online access and downloading. A restricted community of contributors is currently working on increasing the lexical coverage of the resource. Research is underway to extend the knowledge base model to encompass verbal frames.

RÉSUMÉ. Senso Comune est une base de connaissances ouverte de la langue italienne, disponible à travers une plate-forme collaborative sur le Web, dont la construction est en cours. Cette ressource intègre les données lexicographiques provenant à la fois d'utilisateurs et de dictionnaires existants avec une ossature ontologique qui fournit les bases pour une caractérisation formelle de structures sémantiques lexicales (cadres). Un noyau de lemmes italiens de base, qui ont été analysés et classés sémantiquement, est disponible à la fois pour l'accès en ligne et le téléchargement. Une petite communauté de contributeurs travaille actuellement sur l'augmentation de la couverture lexicale de la ressource. Les recherches en cours visent à étendre le modèle de base de connaissances pour inclure les cadres verbaux.

KEY WORDS: electronic dictionaries, linguistic resources, ontology, computational lexicon, lexical semantics, frames, thematic roles.

MOTS-CLÉS: dictionnaires électroniques, ressources linguistiques, ontologie, lexiques computationnels, sémantique lexicale, cadres, rôles thématiques.

1. Introduction

Senso Comune is an on-going project for building an open-knowledge base for the Italian language. Leveraging on Web 2.0, *Senso Comune* is designed as a crowd-sourced initiative that stands on the solid ground of an ontological formalization and well-established lexical resources. The community behind the initiative is growing, and the knowledge base is evolving by integrating user-generated content with existing lexical resources. The ontological backbone provides foundations for a formal characterization of lexical meanings and relational semantic structures, such as verbal frames. *Senso Comune* is an "open-knowledge" project. The lexical resource is available for both online access and downloading[1].

In this paper, we want to present the project, some initial results, and the future directions.

We first illustrate the history and the general goals of the project, its positioning with respect to general linguistic issues, and the state-of-the-art of similar resources. We describe the method to merge crowd-sourced development of the lexical resource and existing dictionaries. We provide some insight of the formal model of the knowledge base, from the perspective of its logical structure and ontological backbone.

The rest of the paper is organized as follows. The next sections describe the motivation and the background. Section 2 provides an overview of the development of the lexical resource. Section 3 describes the formal model of *Senso Comune*. And, finally, section 4 sketches current research and future directions.

1.1. *Project's objectives and history*

In fall 2006, a group of Italian researchers[2] from different disciplines gathered to provide a vision on the role of semantics in information technologies[3]. Among other things, the discussion spotted the lack of open, machine-readable lexical resources for the Italian language. This was seen as one of the major hindering factors for the

---

1. Please visit the project's portal at www.sensocomune.it.
2. Besides the authors of this paper, the group includes Aldo Gangemi, Nicola Guarino, Maurizio Lenzerini, and Malvina Nissim, lead by Tullio De Mauro.

development of intelligent information systems capable of driving business and public services in Italy. Free, high-quality lexical resources such as WordNet contribute to the growth of intelligent information systems in English-speaking countries. Lexical machine-readable resources for Italian – primarily MultiWordnet, EuroWordNet and the follow-up project SIMPLE (see section 1.3.1) – although freely available for research purposes, do not seem to play a similar role in the Italian industry of semantic technologies. From these premises, the group decided to start an open collaborative research initiative, named *Senso Comune* (literally "common sense", but more specifically intended as "common semantic knowledge"). A non-profit association was then established, which holds regular activities and annual workshops since 2007. Beyond the scope of industrial development, the group recognized that an open lexical resource for Italian is a way for collecting and organizing a body of knowledge which is particularly important in a modern country, where the progress in communication technologies increases the pace of linguistic changes.

From the outset, *Senso Comune* was conceived as a linguistic knowledge base rather than a dictionary[4]. It is based on a conceptual apparatus which is not usually present in standard linguistic resources. Besides lemmas and glosses, *Senso Comune* provides a rich semantic qualification. Each sense is mapped to ontological categories and is associated with semantic frames. The starting point to build the knowledge base is a high-quality lexical resource. The legacy resource was digitalized and put into a collaborative platform on the Web, ready to be enriched by a vast (but supervised) community of users. Hence, Tullio De Mauro, who chairs the Association, authorized the use of a small but significant part of his "*vocabolario di base*" (De Mauro, 1980), i.e., the 2,071 most frequent lemmas in Italian. An interdisciplinary, cross-organization team hosted at the Center for Advanced Studies of IBM Italia started designing a representational model (see section 3) and developing the related software tools to accommodate and manage the resource. Fitting the textual dictionary source into the model turned out to be very far from trivial; nonetheless, the Web platform was made available in the 2009, after one year of work.

---

3. IBM Italia Foundation's symposium *La Dimensione Semantica dell'Information Technology* (The Semantic Dimension of Information Technology)*,* Rome, November 27, 2006.
4. As knowledge base we intend here an information system which works under the Open World Assumption, i.e. the assumption that managed data may be incomplete (Russell and Norvig, 2010). As for the implementation concerns, *Senso Comune* currently adopts a standard relational database.

Based on the acquired resource, the second step of the project consisted in classifying 4,586 senses of basic nouns (the most frequent in Italian textual sources) by means of a small set of predefined ontological categories. The work was carried out by undergraduate students under the supervision of the association's researchers (see section 2.2).

In the last year of activity, the development of *Senso Comune* has followed two main tracks. On the one hand, with the aim of providing a large-scale lexical resource, the group focused on how to extend the dictionary to cover thousands of common and less common words. The idea is to blend user contributions with reliable resources in a way that preserves both quality and availability. On the other hand, the group started studying how to extend the model to encompass the kind of lexical knowledge which is not usually represented in traditional lexicography. In particular, a study on verbal frames has been undertaken based on the idea of exploiting the usage examples associated with the sense definitions of the most common verbs included in the dictionary as an empirical base (see section 4).

1.2. *General linguistic orientation of the project*

The *Senso Comune* research group includes linguists, computer scientists, logicians, and ontologists, who look at natural language from different perspectives and with different orientations. The relationship between meanings and reality, that is at the core of lexical semantics and conveys deep philosophical issues, is a largely debated issue. Although the research group members do not share all the assumptions, a common view (synthesized in a *Manifesto*) has been put at the basis of the project. The main assumption is that natural languages exist mainly in their use and the regularities languages show are basically a consequence of social consensus. Since languages serve humans in dealing with the world, ontologies (i.e., theories of the physical and social reality) are essential to characterize such consensus with respect to extra-linguistic entities. In other words, although language is far from being a mere "picture of reality", (hypothetical) pictures of reality are needed to account for lexical semantics, which is where words and entities come into contact. Lexical semantics and ontology, though being different realms, are thus related, and much of the project's specificity is, in fact, the research of a suitable account of such relationship.

The representation of linguistic knowledge in a context-based approach (i.e., dealing with phenomena such as polysemy and ambiguity) is closely related to representations of other kinds of knowledge in the effort to reduce the gap between

the semantic, pragmatic and contextual-encyclopaedic dimensions. The interaction between ontologies, semantics and lexical resources may be established in different ways (Prévot *et al.*, 2010). In our first experiment, we chose to mark linguistic data with concepts of a general formal ontology[5].

Ontologies represent an important bridge between Knowledge Representation and Computational Lexical Semantics, and form a continuum with semantic lexicons[6] (Lenci *et al.*, 2002). The most relevant areas of interest in this context are the Semantic Web and Human-Language Technologies: they converge in the task of pinpointing knowledge contents, although focusing on two different dimensions, namely ontological and linguistic structures. Computational ontologies and lexicons aim at digging out the basic elements of a given semantic space (domain-dependent or general), characterizing the different relations holding among them. Nevertheless, they differ with respect to some general aspects: the polymorphic nature of lexical knowledge cannot be straight off related to ontological categories and relations[7]; polysemy refers to a genuine lexical phenomenon which is generally absent in well-formed ontologies; the formal features of computational lexicons are far from being easily coded in a logic-based language[8]. Since the early 80's, there's been a huge debate in the scientific community on whether the categorical structures of computational lexicons could be acknowledged as ontologies or not (see e.g., Poesio, 2005, for a survey of the issue). The general approach we adopt in *Senso Comune* is to integrate the two dimensions, with no attempt of reducing one to the other.

As far as the ontological layer of *Senso Comune* is concerned, the reader can refer to section 3. In the following sections, we quickly survey two of the most important state-of-the-art computational lexicons, i.e., WordNet and FrameNet, providing the general conceptual framework in which *Senso Comune* is rooted.

---

5. In this respect, our approach is essentially different from OntoNotes (Pradhan *et al.*, 2007), where multi-lingual corpora have been annotated with shallow semantic features based on the Omega ontology. Omega contains "no formal concept definitions and only relatively few interconnections" (Prévot *et al.*, 2010) while *Senso Comune*, conversely, is explicitly grounded on a formal model (see section 3).
6. Here "semantic lexicon" and "computational lexicon" are used as synonyms.
7. A lexicon, by definition, will omit any reference to ontological categories that are not lexicalized in a language.
8. Concerning this last point, for example, a "task force has been created by the W3C Consortium to port WordNet into OWL http://www.w3.org/2001/sw/BestPractices/WNET/tf).

1.3. *Computational lexicons:* **Senso Comune** *vs.* **WordNets and FrameNets**

WordNet was developed in Princeton University under the direction of George A. Miller. Christiane Fellbaum, the principal investigator of the project, describes it as "a semantic dictionary that was designed as a network, partly because representing words and concepts as an interrelated system seems to be consistent with evidence for the way speakers organize their mental lexicons" (Fellbaum, 1998, p. 7).

WordNet is constituted by *synsets* (lexical concepts), namely set of synonym terms – i.e., life form, organism, being, living thing. The idea of representing world knowledge through a semantic network (whose nodes are *synsets*, and whose arcs are fundamental semantic relations[9]) has been characterizing WordNet development since 1985. Over the years, lexicographers have incrementally populated the resource (from the 37,409 *synsets* in the 1989 to about 120,000 *synsets* of the most recent releases), and substantial improvements of the entire WordNet architecture, aimed at facilitating hierarchical organization and computational tractability (accordingly, some OWL-based implementations have been recently released[10]). WordNet covers several domains, namely groups of homogeneous terms referring to the same topic (art, geography, aeronautics, sport, politics, biology, medicine, etc.). In recent years, there have been interesting and fruitful attempts to annotate WordNet with domain/topical information in order to improve the overall accessibility to the dense semantic database[11]. WordNets have been and are being constructed in dozens of languages. Besides the EuroWordNet project that built WordNets for eight European languages, BalkaNet project[12], encompassing six languages, and PersiaNet[13] have been developed. In addition, WordNets are being constructed in Asia and South America[14]. It's also worthwhile to mention the SIMPLE project (Lenci *et al.*, 2000), an evolution of the EuroWordNet project, which implements Pustejovsky's qualia roles (Pustejovsky, 1995).

WordNet has been often referred to as a lexical ontology or – at least – as containing ontological information: although *synsets* can be definitely conceived as lexical counterparts of "ontological categories", WordNet-like resources do not rely on any explicit logical infrastructure. *Senso Comune* has borrowed from WordNet

---

9. Hyponymy, antonymy, troponymy, causality, similarity, etc.

10. e.g., http://www.w3.org/TR/wordnet-rdf/

11. See for example: http://multiwordnet.itc.it/english/home

12. http://www.ceid.upatras.gr/Balkanet/

13. http://persianet.us/

14. For an updated list of WordNet projects see: http://www.globalwordnet.org/

many basic intuitions about lexical ontology. However, *Senso Comune* differs from WordNet in many respects. Instead of focusing on synonymy and hyponymy relations with the aim of bringing out the conceptual structure behind the lexicon, *Senso Comune* adopts a set of *a priori* ontological distinctions, to identify the ontological commitments behind each sense (see sections 2.3 and 3.2).

A semantic lexicon can be structured from a different perspective, focusing on frames instead of *synset*s, as in the case of FrameNet (Ruppenhofer *et al.*, 2005). In the AI tradition, frames are data structures for representing a stereotyped situation, like "in a living room", or "going to a child's birthday party". Minsky describes frames as carrying "several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed" (Minsky, 1997).

FrameNet aims at providing a lexical account of this kind of "schematic representations of situations". Developed at Berkeley University and based on Fillmore's frame semantics (Fillmore, 1968), FrameNet aims at documenting "the range of semantic and syntactic combinatorial possibilities (valences) of each word in each of its senses"[15] through corpus-based annotation. Let's see a sketchy example. If you point to the discussion frame, namely an abstraction of a state of affairs where discussants talk about something in a given place at a given time, you will find several lexical instances in FrameNet (generically called "lexical units" – LUs) of different roles (or frame elements – FEs): i.e., the nouns "president" and "advisor" instantiate the interlocutor role in the frame discussion[16]. In principle, the same LU may belong to distinct frames, thus instantiating different roles: the noun "president", for example, also instantiates the people frame. FrameNet contains about 12,000 LUs in about 1,000 frames (instantiated in about 150,000 annotated sentences). As with WordNet, new projects are under development to yield FrameNet-based computational lexicons for other languages: SALSA project in Germany[17], Spanish FrameNet[18], Japanese FrameNet[19], and domain specific resources like the Soccer FrameNet[20].

---

15. http://framenet.icsi.berkeley.edu/.

16. In this context, an example of annotated sentence is: "The president debated with his top advisor".

17. http://www.coli.uni-saarland.de/projects/salsa/.

18. http://gemini.uab.es:9080/SFNsite.

19. http://jfn.st.hc.keio.ac.jp/.

20. http://www.kicktionary.de/.

Senso Comune's model is being extended to encompass verbal frames (see section 4), which will make it comparable to existing FrameNet-like resources. However, in general, existing FrameNets don't supply any formal characterization of the relations between frames (as well as any specification of the roles embedded in a single frame)[21], which will be a key feature of our approach.

2. Steps in *Senso Comune* development

The *Senso Comune* knowledge base is acquired and maintained by exploiting both lexicographic resources and human contributions, according to the formal model described in the next section. The idea is to allow the resource to be aligned to changing linguistic practice, while keeping the high quality of traditional lexicographical resources.

The whole knowledge base counts about 130,000 lemmas with 240,000 senses, which have been acquired from a number of different sources, as well as users' contributions. At the time being, the general distribution of the knowledge base is limited to a core of 2,071 lemmas of basic Italian, which has been made available by Prof. Tullio De Mauro and elaborated by the community. This resource is available for download in a specific XML format under a Creative Commons Attribution-Non Commercial-Share Alike 3.0 License. The possibility of making available other sections of the resource, including technical lemmas, is being investigated under the profile of copyright protection.

Researchers contributing to the project as well as selected contributors (e.g., students involved in annotation experiments) have access to the whole knowledge base. Qualification as contributor is granted to associates upon specific requests, and subject to the evaluation of the Association's Scientific Committee. The current community counts about 200 people, of which about 40 are active contributors.

2.1. *The acquisition from traditional sources*

To provide the base for the construction of the new resource, the project started from the digitization of a small portion of a paper-based dictionary, namely the

---

21. Some work to enhance Berkeley FrameNet has been done recently by Ovchinnikova, Vieu, Oltramari, 2010.

Grande Dizionario Italiano dell'Uso[22] (GRADIT). At first, this consisted in selecting lemmas marked as "fundamental" (lexemes covering about 90% of all spoken and written texts in Italian). The selection of fundamental vocabulary guarantees maximum text coverage, but at the same time provides complex polysemic lemmas to be processed. In fact, fundamental lemmas are characterized by a large plurality of word senses, often overlapping, which include the most general and abstract meanings. As a second step, the original lemmas associated with glosses, grammatical notes, usage tags and examples were parsed to fit the formal lexicographical model of the knowledge base. This process required looking at lexicographic entries in a formal way, and forced a new reflection on meta-features and norms common to traditional lexicography, that need to be revised when translated into a formal database structure. This is particularly relevant when different kinds of linguistic information sources are to be integrated, as in the case of *Senso Comune*.

The GRADIT dictionary comes with a rich and detailed lexicographic apparatus (De Mauro, 1999) which has driven the development of the *Senso Comune* lexical model. Unfortunately, such formal apparatus is not necessarily reflected in the typographical renderings of paper dictionaries, which are usually conceived for human consumption, thus being often very far from machine-readable formats. This mismatch caused non-trivial problems when parsing the paper dictionary. In many cases, for instance, discriminating among usage examples and sub-senses is implicitly left to the reader, based on evidences which are hard for a machine to analyze. Also, errors or inconsistencies in the paper edition, although negligible by humans, caused rule-based parsing to give unpredictable results. For these reasons, developing and testing the parsing machinery to feed the *Senso Comune* knowledge base has been quite a difficult process. Anyway, users are progressively amending the content, and fundamental lemmas are now free of errors.

2.2. *User-generated content and its validation*

The general idea behind *Senso Comune* is the integration of scientifically-grounded lexicographic resources with the contribution of user generated content, in a controlled "crowd-sourcing" process. As the Wikipedia experience has shown, collaborative projects over the Web can produce large amounts of data, continuously enriched, amended, discussed by users. The elicitation of linguistic knowledge can benefit of Wiki-like approaches because languages are by their own nature fuzzy,

---

22. De Mauro T., 2000, *Grande Dizionario Italiano dell'Uso*, Torino, UTET.

with loose boundaries in norms, competence and grammaticality perception. The best way to represent this intrinsic variation and looseness is to call on to users to perform and to evaluate their linguistic knowledge.

Starting from the baseline of high-quality lexical data, *Senso Comune* allows the community to validate, enrich, and extend the linguistic knowledge base through a Web platform which has been specifically designed and implemented for this purpose. The platform shares with collaborative Wiki-based dictionaries such as Wiktionary[23] a number of features, such as collaboration (content can be inserted, deleted, or edited by multiple users), and traceability (changes are registered so that an incremental history of the single record is maintained). However, *Senso Comune* differs from Wiki platforms in many significant respects. First of all, *Senso Comune* conforms to a rich and specific data model, where senses and their relationships are fully featured objects. On the contrary, Wiki pages are almost blind to conceptual information associated to lexical units. Not by chance, the mechanism Wiktionary provides for handling sense relationships is cumbersome. Moreover, while Wikis force users to a complex editing syntax, *Senso Comune* provides a neat interactive user interface, which is tailored for handling linguistic content.

The main differentiating feature of *Senso Comune* with respect to Wikis is the content-acquisition policy. Users can access the lexicographic resource online and can be given different writing privileges on the database. Every editing operation is monitored and can be tracked as an atomic operation. Read-write controlled access is granted by authorization and is strictly confined to specific portions of the database (e.g., in the experiment described in section 2.3 below, only ontological category association for word senses was enabled), thus allowing different activities to be performed in a controlled flow.

23. www.wiktionary.org
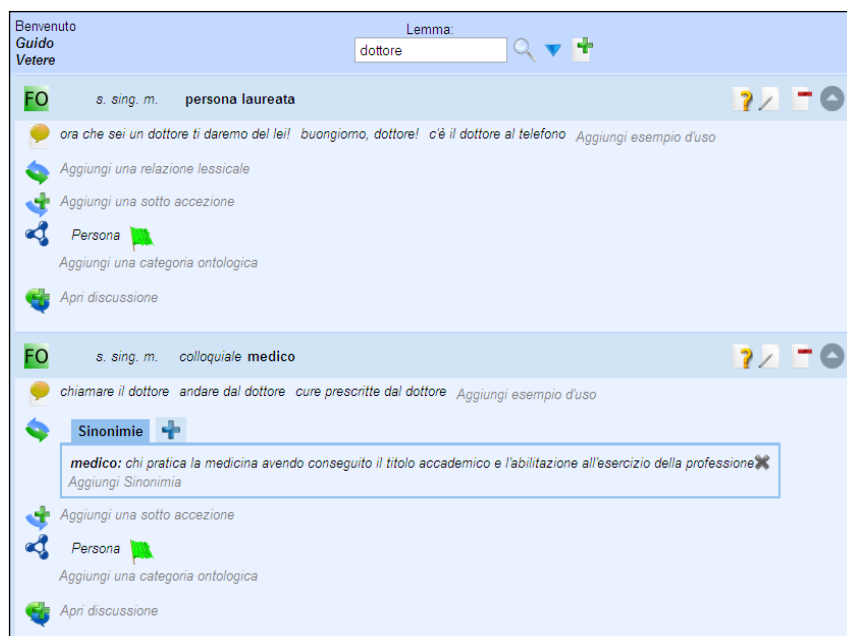
Figure 1. *The User Interface – different senses of the noun "dottore" (doctor), each one with usage instances, lexical relations, ontological classification*

User activities can be guided using different methodologies, as online help and discussion lists, and also by providing tutoring systems that facilitate specific tasks.

### 2.3. *Experimenting noun/word senses ontology tagging*

In order to test advantages and disadvantages of this approach, we adopted a soft start experimenting with a small group of users and introducing various layers of revisions.

The first experimentation has been aimed at observing procedures of association of word senses and ontological categories and to detect and evaluate problems arising during this process. Our primary attempt in this direction has been the association of each of 4,586 word senses (belonging to 1,111 fundamental noun lemmas) to one ontological category and adding a subjective parameter of confidence with which the user could evaluate the degree of doubt of their own

association. The work was carried out by a group of students of Isabella Chiari's graduate computational linguistics class at University of Rome Sapienza.

The procedure was carried out in three phases: I. Primary common sense classification lead by 12 students; II. Revision of the classification (lead by Chiari, Vetere, Oltramari and 4 students) with the additional task of attaching confidence evaluation to the classification using three tags (accepted, controversial, not accepted) and discussion; III. Final revision of consistency in classification actions.

For the annotation of ontological categories, experienced users just pick up an item from a given list. But these categories can be also kept "opaque" to facilitate those who need help to understand ontological commitments behind linguistic senses, as in the case of students involved in the experiment. This has been achieved by using TMEO, a tutoring methodology based on ontological distinctions developed by Alessandro Oltramari at ISTC-CNR's Laboratory of Applied Ontology of Trento. TMEO is a classification system based on broad foundational distinctions from DOLCE-Spray, a simplified version of DOLCE (see section 3.2), which can be implemented with a sequential question answering procedure or a synoptic map (Figure 2). The *Senso Comune* implementation of TMEO helps the user/editor select the most adequate category of the reference ontology as the super-class of the given lexicalized concept: different answer paths lead to different mappings between the lexicon and the (hidden) ontological layer.
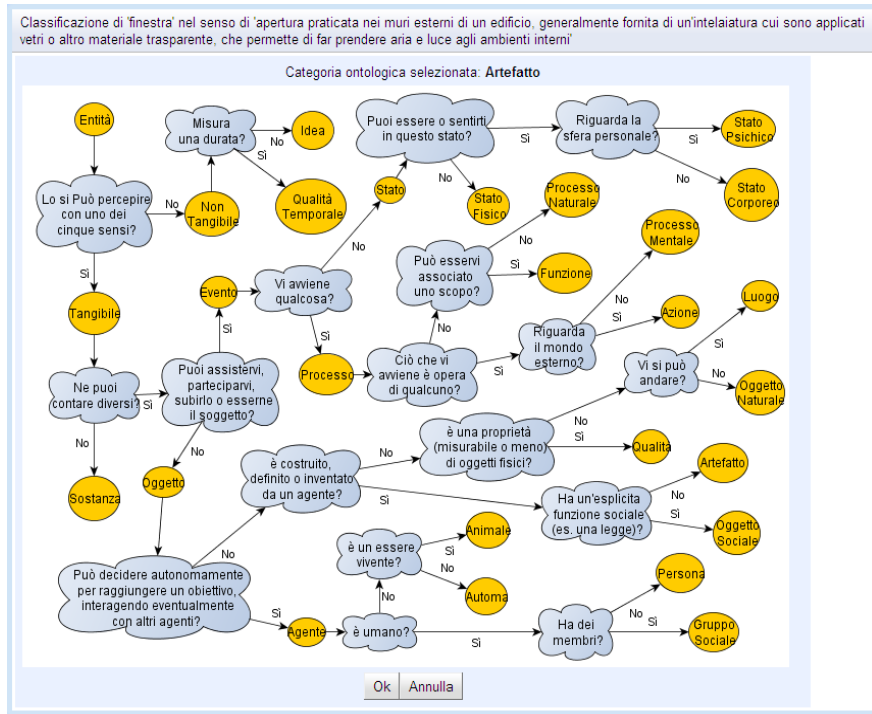
Figure 2. *This conceptual map represents the Q/A mechanism underlying TMEO. DOLCE-Spray categories are represented in yellow circles; state transitions are driven by "yes/no" answers (black arrows) to questions enclosed in blue clouds.*

After six months of work, including supervision, data was analyzed to extract information about word sense distribution in ontological categories, data on categorization problems, and information of variety of ontological classes in the fundamental vocabulary nouns examined.

Ontological category	ws FO	%	Ontological category	ws FO	%
IDEA	689	15,02%	ENTITY	107	2,33%
ARTIFACT	505	11,01%	SUBSTANCE	98	2,14%
PERSON	502	10,95%	SOCIAL GROUP	77	1,68%
QUALITY	433	9,44%	MENTAL PROCESS	74	1,61%
ACTION	413	9,01%	FUNCTION	65	1,42%
NATURAL OBJECT	205	4,47%	PROCESS	65	1,42%
MENTAL STATE	185	4,03%	PHYSICAL STATE	49	1,07%
TEMPORAL QUALITY	184	4,01%	TANGIBLE	46	1,00%
EVENT	172	3,75%	NATURAL PROCESS	42	0,92%
PLACE	170	3,71%	BODILY STATE	33	0,72%
STATE	157	3,42%	ANIMAL	21	0,46%
SOCIAL OBJECT	156	3,40%	AGENT	21	0,46%
OBJECT	107	2,33%	NON-TANGIBLE	10	0,22%

Figure 3. *Fundamental word senses for ontological categories experimented*

The interpretation of data presented in Figure 3 is very complex and involves the consideration of the hierarchical structure of ontological categories and the preference for concrete association exhibited by the experimenters, for details see Chiari *et al.*, 2011. Further problems are posed by the different degrees of confidence in the association process performed and inter-annotator agreement issues: 2,685 (59%) dictionary word senses were steady classified, while 1,537 (33%) caused discussions, and 364 (8%) revealed the ontology to be incomplete or problematic.

Some ontological categories posed more association issues than other (from 68% to 81%). For example, while ANIMAL, PERSON, NATURAL OBJECT, ARTIFACT, SUBSTANCE, ACTION did not pose many confidence questions, a high percentage of discussion and classification instability was raised by categories such as ENTITY, TANGIBLE, NON-TANGIBLE, FUNCTION, OBJECT, STATE, IDEA (Chiari *et al.*, 2011).

As a result of the experiment, the group decided to allow multiple classifications of senses in further experiments, and to broaden the list of ontology concepts. Feedback from actual associations, discussions and confidence degree was further used to make some changes in the ontology and discussing some methodological problems that have arisen during the experimentation.

With a general description of *Senso Comune*'s resource, and its associated platform, we can now give an overview of the formal model of the knowledge base.

3. The Model

3.1. *The formal language*

The choice of Description Logic as the formalism for representing lexical knowledge has been one of the first steps of the project. Description Logic (DL) is a decidable fragment of first-order logic whose associated reasoning procedures are decidable and well understood under the profile of computability (Baader *et al.*, 2007). Properties of DL formalization of lexical semantics have long been discussed from a theoretic standpoint and experimented with in several applications (Franconi, 2007).

The reason of our choice in favour of DL is twofold. On the one hand, DL formal apparatus, specifically tailored for terminological definitions, helps making modelling clear, thus facilitating the specification of the relationship between the lexicographical apparatus and general domain concepts, which underpins lexical semantics. On the other hand, reasoning tasks such as inferring logical inclusion (classification), membership (instance checking) and satisfiability, which DL makes easy, may have interesting applications in computational linguistics. Other benefits include the compatibility with ontology definition languages such as the W3C standard OWL (Ontology Web Language[24]). OWL has well-defined syntax and formal semantics (which are crucial for enabling machine-processing of data), and supports efficient reasoning. Several ontology tools have been developed on top of OWL, such as Protégé-OWL, Swoop, Top Braid Composer. OWL can be split in three different sublanguages: OWL-Lite, syntactically the simplest, is mainly used for expressing class hierarchy and simple constraints; OWL-DL, which provides a rich set of DL constructs, is suitable for most automated reasoning tasks, although it may be problematic for handling large datasets, while OWL-Full, which is the most expressive of three dialects, may lead to undecidability. OWL-2 (released on October 2009) also provides OWL-EL, adopted for large-scale ontologies, OWL-QL specifically designed for being interoperable with database technologies, and OWL-RL, suitable for integration with rule languages.

Limitations of linguistic applications of DL are due to their strictly controlled expressiveness, as well as the binding to first-order logic which, at first glance, does

24. http://www.w3.org/TR/owl-features/.

not support meta-level reasoning. However, while studies on higher-order DL extensions proceed (De Giacomo *et al.*, 2010), ontology definition languages such as OWL-2 provide rich annotation mechanisms that allow syntactic representations of meta-properties, which can be dealt with *ad-hoc* reasoning procedures according to the needs of specific applications.

As mentioned above, DL is well suited for reasoning on classifications and instance checking, that is, calculating whether a class is included in another based on their descriptions, and calculating whether an individual belongs to a class based on its properties, respectively. This appears to be promising when dealing with semantic frames associated to verbs. If semantic frames were given suitable DL representations, discovering generality hierarchies among them would be feasible with current, freely available, reasoning tools. Also, matching semantic frames with linguistic evidences could be reduced, to some extent, to the task of checking whether linguistic tokens fall into formally-defined semantic classes, which appears to be reducible to instance checking.

Even if the current version of *Senso Comune* limits automatic reasoning to formal properties of lexical relations, we anticipate that many of the planned activities (see section 4) could benefit from the highly-formalized declarative form in which linguistic data is collected. Moreover, the entire *Senso Comune* content can be easily exported in a standard knowledge base form, to be dealt with a variety of ontology reasoners, including open source ones.

3.2. *Main model features*

Senso Comune's model is specified in a set of interrelated ontologies comprising a "top level", which contains basic concepts, a "lexical ontology", which models general linguistic and lexicographic structures, and a "frame ontology" providing concepts and axioms for modelling the predicative structure of verbs and nouns[25].

The root of the hierarchy of *Senso Comune* is ENTITY, which is defined as *anything which is identifiable by humans as an object of experience or thought*. The first distinction is among CONCRETE ENTITY, i.e., objects located in definite spatial regions, and ABSTRACT ENTITY, which don't have spatial properties. In the line of Simons (1987), CONCRETE ENTITY is further analyzed in CONTINUANT and OCCURRENT, that is, roughly, entities without temporal parts (e.g., artefacts, animals, substances) and entities with temporal parts (e.g., events, actions, states) respectively.

This top level is similar to that of DOLCE (Gangemi *et al.*, 2002), on which the simplified OWL-Lite version we refer to as DOLCE-Spray, used in the TMEO methodology implemented in the experiment of section 2.3, was based.

DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering)[26] has been explicitly developed in order to address some core cognitive and linguistic features of common sense knowledge. **Figure 4. *Senso Comune*'s "top level" ontology (excerpt)** The basic ontological distinctions are kept: DOLCE's *endurant* and *perdurant* (named OBJECT and EVENT in DOLCE-Spray) match *Senso Comune*'s CONTINUANT and OCCURRENT. The main difference of *Senso Comune*'s top level with respect to DOLCE is the merging of DOLCE's *abstract* and *non-physical-endurant* categories into the *Senso Comune*'s category of ABSTRACT ENTITY.

This move is meant to improve confidence in the annotations with such categories. There was for instance some confusion in the use of DOLCE-Spray's IDEA and

```
ENTITY – anything which is identifiable by humans as an object of experience or thought
        CONCRETE ENTITY – entities with spatial-temporal qualities
                CONTINUANT – concrete entities without temporal parts
                        AGENT – entities which play agentive roles in events
                        GROUP – collection of concrete entities
                                SOCIAL GROUP – intentional collection of
                                humans
                        OBJECT – countable continuant
                                ARTIFACT – objects whose existence roots in
                                agentive processes
                                NATURAL ENTITY – objects whose existence
                                roots in natural processes
                        QUALE – continuants inherent in and existentially depend on
                        other entities
                                SPATIAL LOCATION – physical region occupied
                                by an object
                        SUBSTANCE – non countable continuant
                OCCURRENT – concrete entities with temporal parts
                        PROCESS – events with discrete parts (phases)
                                ACTION – processes initiated by some agent
                        STATE – events without discrete parts
        ABSTRACT ENTITY – entities without spatial qualities
                CHARACTERIZATION – function that maps n-uples of individuals to
                "truth" values
                SOCIAL OBJECT – abstractions accounted within human societies by
                means of linguistic acts
```

Figure 4. *Senso Comune*'s "top level" ontology (excerpt)

25. Ontologies are available at www.sensocomune.org/ontologies.
26. http://www.loa.istc.cnr.it/DOLCE.html.

SOCIAL OBJECT which are now both covered by *Senso Comune*'s ABSTRACT

ENTITY. Among abstract entities, *Senso Comune*'s top level distinguishes CHARACTERIZATION, defined as mapping of *n-uples* of individuals to *truth values*. Individuals belonging to CHARACTERIZATION can be regarded to as "reified concepts", and the irreflexive, antisymmetric relation CHARACTERIZES associates them with the objects they denote. Whether CHARACTERIZATION is formally a *metaclass*, and whether CHARACTERIZES bears the meaning of *set membership* is left opaque in this ontology. Note however that CHARACTERIZATION's subclasses may be restricted to denote instances of specific classes. SOCIAL ROLE, for instance, may be set to characterize only instances of PERSON.

$$\text{SOCIAL ROLE} \sqsubseteq \text{CHARACTERIZATION} \sqcap \forall \text{characterize.PERSON}$$
SOCIAL ROLE is a CHARACTERIZATION which characterizes only PERSONs

Of course, any application which may need to handle instances of CHARACTERIZATION as metaclasses has to be specifically (and carefully) designed. As mentioned above, higher-order Description Logic may provide such applications with formal foundations.

SOCIAL OBJECTs are, in line with Searle (1995), abstractions dependent upon (i.e., constructed and maintained within) human societies with the purpose of characterizing the function of other entities. Indeed, linguistic entities are objects of this kind. To account for them, *Senso Comune* provides a Linguistic Ontology which

EXPRESSION – *the class of all of spoken or written meaningful and combinable linguistic entities*
LINGUISTIC FORM – *characterization of linguistic expressions based on their structure*
      WORD – *linguistic form of atomic expressions*
      PRHASE – *linguistic form of composite expression*
LINGUISTIC PROPERTY – *characterization of linguistic forms*
      MORPHOSYNTAX CLASS – *classes of words, related to their role in phrases*
          NOUN
          VERB
          ...
      MORPHOSYNTAX ATTRIBUTE – *properties of words, related to their classes*
          GENDER
          NUMBER
          …

Figure 5. *Senso Comune* "linguistic ontology" (excerpt)

roots in the joint of CHARACTERIZATION and SOCIAL OBJECT (Figure 5). Central to *Senso Comune* is the notion of MEANING, which is a SOCIAL CHARACTERIZATION of ENTITIES carried by linguistic EXPRESSIONS. Meanings bear lexical relations such as SYNONYMY or HYPONYMY. However, unlike most of the existing lexical semantic models, *Senso Comune* does not give these relations a direct ontological counterpart. In fact, linguistic meanings are entities of their own kind, which are related with the entities they refer to through the CHARACTERIZE first-order relation. Thus, predicating on meanings (e.g., about their relatedness) does not imply predicating on their referents. Specific consistency-checking (or enforcing) procedures can be designed to ensure, for instance, HYPONYMY (of meanings) to be consistent with INCLUSION (of their referred entities' classes), or SYNONYMY to correspond to EQUIVALENCE. Note however that, due to the ontological underspecification of words in the contexts in which they occur, these constraints could be hardly enforced. In fact, meanings referring to what appear as disjoint classes of entities (e.g., ARTIFACT and SOCIAL OBJECT) may overlap in many contexts, as witnessed by the phenomenon of co-predication in systematic polysemy, and thus perceived as lexically related, as illustrated now.

Assigning one or more ontological categories to the meanings of *Senso Comune* has the formal consequence of restricting their characterizations. For instance, classifying the meaning BOOK-1 (*a set of printed sheets*) of the noun "book" as ARTIFACT, corresponds to asserting:

BOOK-1 $\sqsubseteq$ MEANING $\sqcap$ $\forall$ characterize.ARTIFACT
BOOK-1 is a MEANING which characterizes only ARTIFACT

But the meaning of "book" can also refer to the text written, an abstract entity of category SOCIAL OBJECT, as in sentences like *this book on logic is hard going*, and in fact to both the artifact and the social object at once in sentences like *Mary burnt the book on logic*.
By further assigning BOOK-1 the class SOCIAL OBJECT, the assertion above is modified as follows:

BOOK-1 $\sqsubseteq$ MEANING $\sqcap$ $\forall$ characterize.(ARTIFACT•SOCIAL OBJECT)
where ARTIFACT• SOCIAL OBJECT $\equiv$ COMPLEX-TYPE $\sqcap$ =1 member.ARTIFACT $\sqcap$ =1 member.SOCIAL OBJECT
BOOK-1 is a MEANING which characterizes only entities of the complex type ARTIFACT• SOCIAL OBJECT where the category ARTIFACT• SOCIAL OBJECT is a COMPLEX-TYPE whose instances are made up of exactly one ARTIFACT and one SOCIAL OBJECT.

This example illustrates how *Senso Comune,* at the current stage, approaches the complexity of ontological classification, which is at the core of systematic polysemy effects in language (Pustejovsky, 1995). Fully addressing lexical semantics open problems such as *coercion* and *co-predication* is out of the scope of the project at the moment. Nonetheless, the position of *Senso Comune* towards comprehensive theories for such issues is being developed.

3.3. *Data access and reasoning*

Technically, the *Senso Comune* model is a set of OWL-DL ontologies[27] corresponding to the conceptualization outlined in the previous section. Moreover, the model is mapped with a relational scheme, which gives form to a database which can be browsed, queried and populated by users through a Web platform. The conceptual-relational mapping optimizes the knowledge base persistence both for update and query operations, which work under the Open World Assumption. The logical entailments specified by modelling ontologies, if relevant for application purposes, are granted the system implementation. Proposals for transferring the database content in a standard knowledge base manager are under evaluation. The current SQL database ensures efficiency for the most common update and retrieval operations. On the other hand, to suitably implement the required functionalities, queries and application code must be carefully analyzed and designed.

Potentially, managing lexical relationships is one of the most complex tasks among those a lexical knowledge base is expected to support. This is due to formal properties of such relations, such as transitivity, symmetry/antisymmetry, and disjointness, and their interplay. However, *Senso Comune* does not implement transitivity for synonymy and hyponymy, thus facilitating the task of managing lexical relations. This choice is not dictated by technical reasons, but theoretical ones (Cruse, 1986). At the moment, the system ensures synonymy and antonymy to be disjoint, synonymy to be symmetric, and hyponymy\hyperonymy to be antisymmetric and mutually inverse.

The rich structure of *Senso Comune* gives applications the possibility of performing sophisticated conjunctive queries. For instance, one could retrieve all the meanings belonging to a given ontological class whose definition contains a given substring. However, common users (including specialists) are not expected to be able to form such complex queries in a complex query language. Therefore, the Web

27. Ontologies of *Senso Comune* are available at www.sensocomune.org/ontologies.

platform provides a query interface that, at the moment, allows setting lemma, definition and grammatical constraints, as well as retrieving lemmas that the user has commented on.

4. On-going work: semantic role annotation

The main current research activity of *Senso Comune* is directed towards extending the lexical knowledge base to encompass syntactic and semantic verbal frames. The basic idea is to annotate the examples associated with the sense definitions of 740 fundamental verb lemmas (about 4,500 senses) with syntactic and semantic information about frame participants, and to use this information to induce the corresponding verbal frames. In the rest of the section, we first give an overview of the adopted syntactic model, then focus on the annotation schemes under construction.

4.1. *Syntactic annotation*

To represent syntactic structures, we decided to use the extended dependency graph (XDG) (Basili and Zanzotto, 2002). This representation has been proposed as a shared syntactic representation in pipelines of syntactic parsing modules (e.g., tokenization, part-of-speech tagging, chunking [Abney, 1996] verb argument detection, etc.). Merging constituency (Chomsky, 1965) and dependency (Tesnière, 1959) syntactic representations in a single formalism, the XDG formalism has two interesting properties: it hides unnecessary ambiguity in possibly underspecified constituents, and it may represent alternative interpretations in a single graph.

An XDG is defined as a dependency graph whose nodes C are constituents and whose edges D are the grammatical relations among the constituents, i.e., $XDG = (C,D)$ (see Figure 6 for an example). Constituents (i.e., c in C) are classical syntactic trees with explicit syntactic heads, $h(c)$, and potential semantic governors, $gov(c)$ (Pollard and Sag, 1994).
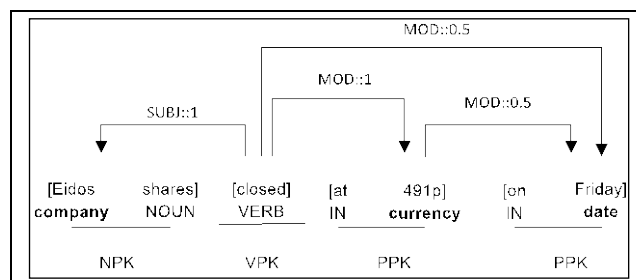
Figure 6. *An eXtended Dependency Graph*

Constituents are chunks that are non-recursive kernel phrases (Abney, 1996), e.g., noun phrases (NPK), verbal phrases (VPK), and prepositional phrases (PPK). A constituent can be either simple (i.e., a preterminal node) or complex (i.e., a syntactic subtree). We represent a constituent as a structure [T ID word_span] where

- T is its type (e.g., NPK, VPK, PPK);

- ID is its unique identifier;

- word_span are the words that are covered.

Dependencies *(h,m,T,plaus)* represent ambiguous relations among a constituent, the head *h*, and one of its modifiers *m* with a type *T* and a plausibility *plaus*. The plausibility reports the ambiguity of the relation. The value ranges between 0 and 1 (not-ambiguous).

In *Senso Comune*, first we will automatically parse verb-usage examples for sense definitions and, then, we will manually correct these syntactic annotations. We will use Chaos (Basili and Zanzotto, 2002) – a robust modular parser that produces interpretations for Italian in the XDG formalism. Syntactic annotation of the dataset of verb-usage instances will drive the subsequent manual activity of semantic annotation. As an example, we report hereafter the interpretations of a sample of sentences associated with the sense definitions of *leggere* (to read) and *annunciare* (to announce) taken from the pilot annotation experiment we run to release the beta version of the annotation scheme. We report constituents along with dependencies among constituents:

(1) *leggere* (to read)–
  — prendere conoscenza del contenuto di uno scritto attraverso la lettura (acquire knowledge of a written work's content through reading)

a. [VPK 1 l.] [NPK 2 un autore (an author)](1,2,VOBJ)
b. [VPK 1 l.] [NPK 2 un libro (a book)](1,2,VOBJ)
c. [VPK 1 l.] [NPK 2 un testo (a text)] [PPK 3 in chiave filologica (from a philological perspective)] (1,2,VOBJ)(1,3,MOD)

(2) *annunciare* (to announce)
— far sapere, rendere noto, comunicare (make known, notify, communicate)
a. [VPK 1 a.] [NPK 2 l'arrivo (the arrival)] [PPK 3 *del volo* (of the flight)] [PPK 4 *da Parigi* (from Paris)] (1,2,VOBJ) (2,3,MOD) (3,4,MOD)
b. [NPK 1 *vi* (to you)] [VPK 2 *annuncio*] [SubC 3 *che* (that)] [VPK 4 *ho ottenuto* (I have gained)] [NPK 5 *la promozione* (a promotion)] (2,1,MOD)(2,4,VSub)(4,3,Sub)(4,5,OBJ)
— riferire il nome di un visitatore (report the name of a visitor)
c. [NPK 1 *il maggiordomo* (the butler)] [VPK 2 *annunciò*] [NPK 3 *la contessa* (the countess)] (2,1,VSUBJ)(2,3,VOBJ)

Parsing the examples associated with the sense definitions is tricky as these examples are generally in the infinite form, do not have a subject, and the verb of the definition is a shortcut, e.g., *l.* instead of *leggere* (to read). Adapting the parser to the specific language is then a precondition to have good results (as in early work [Slator, 1989]).

4.2. *Semantic annotation*

In semantic annotation every frame participant in the parsed sentences will be tagged with its relevant ontological category and thematic role. The basic idea is that, since we assume that there is a principled distinction between the inherent properties of a frame participant and the way it is involved in the event, we want this distinction to be reflected systematically in the annotation scheme. After examining alternative approaches to the representation of semantic role information, particularly VerbNet (VN) and LIRICS (Petukhova and Bunt, 2008) and the on-going attempt to create a unified standard set for the International Standard Initiative (ISO) (Bonial *et al.*, 2011), a set of 25 coarse-grained (high level) semantic roles and their definitions was designed for *Senso Comune*.

In the annotation process, users will be given the sense definitions for each target verb and the associated instances of use. For each frame participant realized in the example, they will be asked to attach a semantic role. For this aim, we will provide them with a list of semantic roles with definitions and examples, together

with decision trees for some of the semantic roles with rather subtler differences, to help distinguish confusing cases.

Participants will also be required to attach an ontological category to each nominal frame participant. As in the previous experiment of "ontologization" of noun senses (see section 2.3), the TMEO method will be used to help them selecting the right category in a new simplified ontology based on *Senso Comune*'s top level (see section 3.2). Drawing on the results of the previous experiment, we will allow multiple ontological classifications, that is, we will allow the users to annotate more than one ontological category on the same argument filler. Such data will be used to investigate further how to handle systematic polysemy in the resource. In the examples (3) and (4) below, we report the results of the pilot semantic annotation performed on the examples presented in section 4.1. Where needed, we add a placeholder for the missing logical subject. We have two ways for indicating these missing subjects: (1) $\varnothing$ is used when the verb is in the infinite form; (2) 0 is used when the verb is in a finite form and the subject is not expressed.

(3) *leggere* (to read)
    — prendere conoscenza del contenuto di uno scritto attraverso la lettura
    (acquire knowledge of a written work's content through reading)
    a. [$\varnothing$ AGENT / person] l. [un libro (a book) THEME / idea+artifact]
    b. [$\varnothing$ AGENT / person] l. [un autore (an author) SOURCE / person]
    c. [$\varnothing$ AGENT / person] l. [un testo (a text) THEME / idea] [in chiave filologica
    (from a philological perspective) MANNER]

(4) *annunciare* (to announce)
    — far sapere, rendere noto, comunicare (make known, notify, communicate)
    a. [$\varnothing$ AGENT / person] a. [l'arrivo del volo da Parigi (the arrival of the flight
    from Paris) THEME / process]
    b. [0 AGENT / person] [vi GOAL / person] annuncio [che ho ottenuto la
    promozione (that I have gained a promotion) THEME / process]
    — riferire il nome di un visitatore (report the name of a visitor)
    c. [il maggiordomo (the butler) AGENT / person] annunciò [la contessa (the
    countess) THEME / person]

Annotated data can be used to conduct an extensive study of the interplay between thematic role information and ontological constraints associated with the participants in a frame; to refine the ontologization of nouns senses in *Senso Comune* by assigning ontological classes to nouns in predicative context instead of nouns in isolation; to investigate systematic polysemy effects in nominal semantics on a quantitative basis. Our long-term goal is to build a rich ontology of Verb Types within the *Senso Comune* infrastructure, informed by empirical data.

5. Conclusion

In this paper we have described *Senso Comune*, a project for building an open-knowledge base for Italian. As a lexical resource, *Senso Comune* is based on an innovative model which relates linguistic phenomena with extra linguistic realities, while keeping language and ontology clearly separated. A rich Web platform allows a community of linguists to feed a knowledge base built around a core of 2,071 basic Italian lemmas, with the aim of opening the resource to the whole community of Italian speakers. Started in winter 2007, the project has taken advantage of many findings coming from significant experiences such as WordNet and FrameNet, and has taken into account the most recent advances in lexical semantics. The current research focuses on acquiring relevant amounts of data on verbal frames, by involving a large community of trained contributors.

Although it is a relatively new initiative, *Senso Comune* has already produced interesting results. Classifying dictionary's fundamental meanings by means of formal ontological categories, based on common sense intuitions, has shown to pose some challenges, but did not result an impossible task. This activity can lead to identify the most basic ontological commitments that characterize lexical semantics, which can foster advances in modelling linguistic senses and lexical ambiguity. Also, *Senso Comune* has shown that users' linguistic elicitation can go beyond definitions and grammar specifications to reach the level of formal semantics. We deem interfacing ontologies and computational lexicons as a key move to pursue for next generation knowledge systems, and believe that *Senso Comune* is an innovative experience in this direction.

Finally, we believe that linguistic knowledge is a common good, and should thus be shared by the entire community, like the language itself. However, as a matter of fact, high-quality linguistic resources are often protected by copyright. The rich structure of *Senso Comune* knowledge base makes information coming from these resources usable for research purposes, and helps granting access to freely available

content in a selective way. We hope that this will be a step towards a greater availability of open-source lexical resources.

6. Bibliography

Abney S., 1996. *Part-of-Speech Tagging and Partial Parsing*. In Bloothooft G. Church K., Young S., editors, *Corpus-Based Methods in Language and Speech*. Kluwer academic publishers, Dordrecht, 1991.

Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P.F., eds, *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2nd edition, 2007.

Basili R., Zanzotto F. M., *Parsing Engineering and Empirical Robustness*, Natural Language Engineering, 2002.

Bonial, C., Brown S.W., Corvey W., Petukhova V., Palmer M., Bunt H., *An Exploratory Comparison of Thematic Roles in VerbNet and LIRICS*. In Proceedings of the 6th Joint ISO – ACL SIGSEM Workshop on Interoperable Semantic Annotation, 2011.

Chiari I., Oltramari A., Vetere G., *Di Cosa Parliamo Quando Parliamo Fondamentale?*, in Ferreri S., editor. Atti del Convegno della Società di linguistica italiana, Roma, Bulzoni, 2011, p. 221-236.

Chomsky N., *Aspects of the Theory of Syntax*. Cambridge, The MIT Press, 1965.

Cruse D., *Lexical Semantics.* Cambridge University Press, Cambridge, England, 1986.

De Giacomo G., Lenzerini M., Rosati R., *On Higher-Order Description Logics*, 2009, available at: http://www.dis.uniroma1.it/~degiacom/papers/2009/DL09.pdf.

De Mauro T., *Guida all'Uso delle Parole*, Editori Riuniti, Roma, 1980.

De Mauro T., *Grande Dizionario Italiano dell'Uso*. Torino, UTET, 1999-2000.

Dowty, D., *Thematic Proto-Roles and Argument Selection*. In *Language*, 67, 1991, p. 547-619.

Fellbaum C., editor. *WordNet – An Electronic Lexical Database*. Cambridge, Massachusetts. MIT Press, 1998.

Fillmore J.C., *The Case for the Case.* In Bach E. and Harms T., editors, *Universals in Linguistic Theory,* Rinehart and Winston, New York, 1968.

Franconi E., *Natural Language Processing*. In Baader F. *et al.*, editors, *The Description Logic Handbook,* cit. 2007.

Gangemi A., Guarino N., Masolo C., Oltramari A., Schneider L., *Sweetening Ontologies with DOLCE*. Proceedings of EKAW, 2002, p. 166-181.

Hanks P., Pustejovsky J., *A Pattern Dictionary for Natural Language Processing*, in *Revue française de linguistique appliquée*, 10/2, 2005, p. 63-82.

Hirst G. *Ontology and the Lexicon.* In Staab S. and Studer R., editors, *Handbook on Ontologies*, Springer, Berlin, 2004, p. 209-229.

Lenci A., Bel N., Busa F., Calzolari N., Gola E., Monachini M., Ogonowski A., Peters I., Peters W., Ruimy N., Villegas M., Zampolli A., *SIMPLE: a General Framework for the Development of Multilingual Lexicons*. In *International Journal of Lexicography*, 13(4), 2000, p. 249-263.

Lenci A., Calzolari N., Zampolli A., *From Text to Content: Computational Lexicons and the Semantic Web*. In 18th National Conference on Artificial Intelligence; AAAI Workshop, "Semantic Web Meets Language Resources", Edmonton, Alberta, Canada, 2002.

Minsky M., *A Framework for Representing Knowledge*. In Winston P., editor, *Mind Design*, MIT Press, 1997, p. 111-142.

Oltramari A., *Tutoring Methodology for the Enrichment of Ontologies* "Cahiers de lexicologie". Forthcoming (December 2011).

Ovchinnikova E., Vieu L., Oltramari A., Borgo S., and Alexandrov T., *Data-Driven and Ontological Analysis of FrameNet for Natural Language Reasoning*. In Calzolari N., Choukri K., Maegaard B., Mariani J., Odijk J., Piperidis S., Rosner M., and Tapias D., editors, Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, 2010. European Language Resources Association (ELRA).

Petukhova,V., Bunt H., *LIRICS Semantic Role Annotation: Design and Evaluation of a Set of Data Categories*. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco, May, p. 28-30.

Poesio M., *Domain Modelling and NLP: Formal Ontologies? Lexica? Or a Bit of Both?,* Applied Ontology (1), 2005, p. 27-33.

Pollard C., Sag I.A., *Head-Driven Phrase Structured Grammar*. Chicago CSLI, Stanford, 1994.

Pradhan S., Hovy E.H., Marcus M., Palmer M., Ramshaw L., and Weischedel R., *OntoNotes: A Unified Relational Semantic Representation*. In Proceedings of the 1st IEEE International Conference on Semantic Computing (ICSC-07), Irvine, CA.

Prévot L., Huang C., Calzolari N., Gangemi A., Lenci A., Oltramari A., *Ontology and the Lexicon: A Multi-Disciplinary Perspective*. In Huang C., Calzolari N., Gangemi A., Lenci A., Oltramari A., and Prévot L., eds, Cambridge University Press, 2010, p. 3-24.

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

Ruppenhofer J., Ellsworth M., Petruck M.R.L., Johnson C.R., *FrameNet: Theory and Practice*, 2005. Available at: http://framenet.icsi.berkeley.edu/.

Russell S., Norvig P., *Artificial Intelligence, a Modern Approach,* Prentice Hall, 2010.

Searle J., *The Construction of Social Reality*, The Free Press, New York, 1995.

Simons P., *Parts: a Study in Ontology*, Clarendon Press, Oxford, 1987.

Slator B.M., *Extracting Lexical Knowledge from Dictionary Text*. In *Knowledge Acquisition*, volume 1, issue 1, p. 89-112, March 1989.

Tesnière L., *Éléments de syntaxe structurale*, Klincksiek, Paris, France, 1959.