An open knowledge base for Italian language in a collaborative perspective

Isabella Chiari, Aldo Gangemi, Elisabetta Jezek, Alessandro Oltramari, Guido Vetere, Laure Vieu

An open knowledge base for Italian language in a collaborative perspective

Isabella Chiari
Sapienza Università di Roma
p.le Aldo Moro 5
00185 Roma
+390696040351
isabella.chiari@uniroma1.it

Aldo Gangemi
Université Paris 13 Sorbonne Paris
LIPN, CNRS (UMR 7030)
Villetaneuse, France
+390644161535
aldo.gangemi@cnr.it

Elisabetta Jezek
Università di Pavia
Dipartimento di Studi Umanistici
Strada Nuova 65 27100 Pavia
+39-382-984391
jezek@unipv.it

Alessandro Oltramari
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA 15213
+1-412-268-6284
aoltrama@andrew.cmu.edu

Guido Vetere
IBM Center for Advanced Studies, Italy
Via Sciangai 53, 00144, Rome, Italy
+39659662137
gvetere@it.ibm.com

Laure Vieu
IRIT-CNRS, Université Toulouse III
118, route de Narbonne
31062 Toulouse cedex 9
+33 (0)561557713
vieu@irit.fr

ABSTRACT

In this paper, we describe the architecture and collaborative approach adopted for the development of Senso Comune (SC), an open knowledge base for the Italian language that combines lexicographic, linguistic, ontological and textual data in a web-based collaborative platform. Senso Comune aims at integrating lexicographic data from users and legacy resources with a formally characterized ontology grounded on lexical semantic structures. Senso Comune was conceived as a linguistic knowledge base rather than a dictionary. It is actually based on a conceptual apparatus that is unusual in typical linguistic resources:, each sense is mapped to ontological categories, and is associated with semantic frames. A wiki provides a collaborative platform for editing a basic lexicon of Italian with different access and annotation privileges that dynamically enables crowd-sourced annotations. Senso Comune proposes different levels of interrelated representation layers that differently exploit collaborative annotation.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content analysis and Indexing – *Abstracting methods, Dictionaries, Indexing methods, Linguistic processing, Thesauruses.*

Keywords

Electronic dictionaries, linguistic resources, ontology, computational lexicon, lexical semantics, frames, thematic roles.

1. INTRODUCTION

The notion of document has undergone significant changes in the last ten years going toward an integrated approach that includes all kinds of data (text and multimedia), and giving strong relevance to the way textual and non textual data are organized, annotated, interlinked and managed in open and closed environments. The area of computational lexicography and that of computer-aided traditional lexicography has been deeply affected by recent changes in IT and by the possibilities of integrating different representations of linguistic data in a single environment giving the user (human and machine) different interpretative perspectives over the same data sources.

Particularly successful has been the case of WordNet [1] and projects inspired by its approach where semantic networks have been exploited both in the direction of computational application and of a more formal representation of hierarchical relationships between senses and synsets, as in WordNet RDF- and OWL-based implementations [2]. WordNet has also often been considered as a lexical ontology or at least as containing ontological information. Nevertheless, although synsets can be conceived as lexically grounded counterparts of ontological categories, WordNet-like resources do not rely on an explicit formal semantics. A similar claim can be made for other lexical resources such as FrameNet [18] and VerbNet [19], which focus respectively on frame- and verb-based lexical semantics, and have been ported to RDF and OWL (cf. [16] for FrameNet-OWL). Porting lexical data into semantic web knowledge bases is relevant in order to provide a lexical grounding to the fast growing linked data cloud.

In the Multilingual Semantic Web,[1] a typical assumption is that lexical entities are individuals in a (lexical) domain of discourse, while specific applications can possibly attempt to reuse lexical entities with a domain-oriented semantics, e.g. assuming synsets as classes or semantic roles as binary relations. Similarly to what is described in [2][16][17], and in the currently undergoing standard proposal from the OntoLex W3C Community Group,[2] Senso Comune provides a direct link between semantic entities from lexica, like senses, synsets, frames, roles, etc., and semantic entities from ontologies.

The emerging trend for modelling lexical semantics in semantic applications is to allow a meaning or sense layer between linguistic expressions and entities (individuals, classes, relations) from ontologies. For example, the word dog from WordNet can be linked to its sense word-sense-dog-1, which on its turn can be linked to a `:Dog` class from an ontology. Similarly, the lexical unit heal from FrameNet can be linked to its Cure frame, which on its

[1] Cf. the Multilingual Semantic Web Workshop, http://ceur-ws.org/Vol-936/

[2] http://www.w3.org/community/ontolex/

turn can be linked to a :Cure class (or relation) from an ontology. This modelling pattern is in line with classical semiotics, and assumes an intensional semantics for lexical meanings vs. an extensional semantics for ontology entities. The links however can be navigated to reuse intensional meanings as extensional when required. While there are two Italian wordnets available (e.g. MultiWordNet [3] and ItalWordNet [4], and there have been attempts at automatically inducing an Italian FrameNet [5, 6], there is as yet no VerbNet-like resource for Italian.

The present contribution will provide an overview of the different modules and data that are joined in the resource, discussing the integration of perspectives and annotation rationales and also illustrating how collaborative annotation has been introduced in the overall project. We will offer a general introduction to the Senso Comune project, its purpose, features and current state of development (§2). The central part of the contribution explains how the different layers are integrated in the general architecture of SC and how the architecture is represented in the platform and in the downloadable resource (§3). We will finally exemplify and discuss how collaborative annotation is currently used for tagging different kind of data and what are the tools developed in order to support annotation (§4).

2. SENSO COMUNE: THE PROJECT

Senso Comune started in 2006 as a collective initiative promoted by a group of researchers[3], lead by linguist Tullio De Mauro, interested in the development of an open lexical resource for the Italian language. Senso Comune has been established as a non-profit association and holds regular activities and annual workshops since 2007.

Senso Comune (literally "common sense", but more specifically intended as "common semantic knowledge") is not conceived merely as an electronic dictionary but as a knowledge base where different kinds of data are integrated, connected and annotated.

One of the main objectives of Senso Comune is in the internal formalized representation of linguistic knowledge such as lexical and morphological information, semantic specifications through ontologies, and thematic roles and frames. Senso Comune is devoted to the distribution of linguistic data in an open and standardized form. The resource is available for download in a specific XML format under a Creative Commons Attribution-Non Commercial-Share Alike 3.0 License.

The starting point of the resource was the so-called fundamental vocabulary of Italian language [7], containing the top-ranked 2,000 lemmas (and about 13,000 senses) extracted from frequency lists of written and spoken Italian and covering about 90% of the occurrences of any written or spoken text. The digitalized version of the fundamental vocabulary was then remodelled in order to fit into a formal representation (meta-model) specifically designed for integrating lexical and ontological information [8, 9]; finally, the resource was shared online in 2009, hosted by the Center for Advanced Studies of IBM Italia.

The online platform allows controlled user contributions, with the aim of providing a reliable, large-scale resource that preserves both quality and availability. Through the platform, the resource was later enriched with the classification of 4,586 senses of basic nouns (1,111 nouns) by means of a small set of predefined ontological categories. This work was carried out collaboratively by undergraduate students under the supervision of researchers. The resource has been recently integrated with the Italian version of MultiWordNet [3], whose content has been arranged to fit in the Senso Comune meta-model. The pairwise alignment of MultiWordNet senses to Senso Comune's, relatively to overlapping lemmas, is currently being studied, and semi-automatic alignment methods are being evaluated.

Senso Comune correlates linguistic meanings and ontological concepts while keeping those levels conceptually and architecturally separate (see Figure 1): meanings are not modelled as generic ontological concepts, but rather, as social abstractions that abridge lexemes and their ontological relata together. In these regards, our approach rests on the duality of language and ontology rather than embracing some form of monism à la Gadamer [20] and reducing one to the other. In fact, natural language texts do not reveal just the core relations between entities in a specific domain, e.g., a war and its triggering events, but typically embed more complex aspects of the human communication framework, including uncertainty, cognitive and metacognitive dimensions, emotional gradations, cultural imprinting and social interactions. If ontologies are powerful tools for shaping this stranded "linguistic matter" into machine-readable "semantic objects", they also carve out those dynamic aspects of meaning along the process, turning human linguistic contents into a set of properties that satisfy some computational constraints. By decoupling the linguistic and ontological level in Senso Comune, conflicts and inconsistencies in user inputs do not directly affect the ontology and maintain their characteristic richness and variety; additionally, there's room for introducing automatic, semi-automatic, or manual procedures to map linguistic senses to their ontological counterparts.

Current research is focused on ways to extend the model to encompass the kind of lexical knowledge that is not usually represented in traditional lexicography. More specifically, the integration of verbal frames in the resource with annotation of usage examples associated with the sense definitions of the most common verbs included in the dictionary as an empirical base was started (see section 3.4).

2.1 Senso Comune: The Platform

A Web based platform to make Senso Comune available, and to manage and develop the resource has been designed and developed at the IBM Center for Advanced Studies of Rome. The platform shares with lexicon oriented "wikis" like Wiktionary[4] a number of key features. However, it support the collaboration on lexical-ontological resources in a more specific way. Primarily, instead of using generic wiki structures and formatting (which makes Wiktionary somewhat difficult to use), Senso Comune platform (www.sensocomune.it) is tailored on a specific data model, and provides contributors with a rich and easy user interface (Figure 1). On the one hand, this frees users from dealing with complex editing and structuring conventions; on the other hand, it reduces the risk of injecting errors in the lexical resource, due platform's technicalities.

Internally, the platform manages an information system structured around a set of Java modules, which link the one another by way of programming interfaces (API). Application modules are arranged in two architectural layers: the front-end (which handles users' interactions) and the back-end, which

[3] Among those who contributed to the design and development of Senso Comune we must mention: Tullio De Mauro (President), Guido Vetere (vice-President), Diego Calvanese, Isabella Chiari, Aldo Gangemi, Nicola Guarino, Elisabetta Jezek, Maurizio Lenzerini, Malvina Nissim, Alessandro Oltramari, Laure Vieu, Fabio Massimo Zanzotto.

[4] http://www.wiktionary.org/

manages data accesses and transactions. The database where the resource is stored has a schema which efficiently supports complex queries, such as, for instance, looking for all the lexical entries whose meanings are classified with a given ontology concept. Also, integrity of data records is preserved, based on RDBMS standard functionalities. An Open Source object-relational mapping utility (Hibernate) ensures a smooth alignment of data records with programming structures.

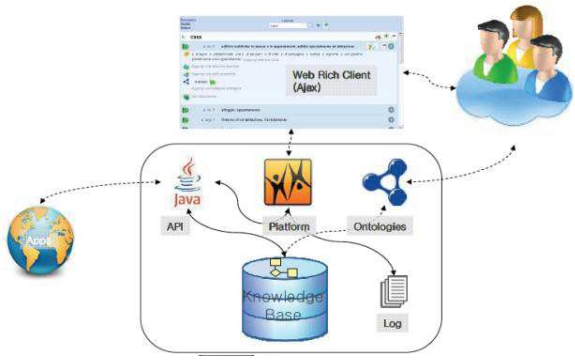

Figure 1. Senso Comune: the platform

2.2 The Model

Senso Comune's model is specified in a set of "networked" ontologies [10] comprising a top level module, which contains basic concepts and relations, a lexical module, which models general linguistic and lexicographic structures, and a frame module providing concepts and axioms for modelling the predicative structure of verbs and nouns. These ontologies have been given an OWL specification, thus leveraging the expressive power of Description Logic (DL) [11]. This way, the resource benefits of a high level of standardization, as well as of the availability of open source tools. Embracing a DL framework also allows for using well understood reasoning procedures, such as checking logical consistency and coherence, evaluating class membership (instance classification), and executing logical conjunctive queries (query answering). These functionalities are provided by open source inference engines that enhance the resource's exploitability. For the sake of efficiency over large lexicographic data sets, the platform stores the resource in a specifically designed relational database, as illustrated in the previous paragraph. To ease the development of ontology-based access to the lexical resource in the future, however, the database schema strictly conforms to the ontology.

We will not go into further detail here on the specific features of the meta-model, previously described in [8, 9].

2.3 Experimenting word sense and ontological category mapping

Meanings from De Mauro's core Italian lexicon have been clustered and classified according to concepts belonging to Senso Comune's model, through a supervised process. To enrich the knowledge base, though, language users have been given access to the lexical level only. This access restriction produces an epistemological spread between ontological and linguistic dimensions, but this gap is a necessary requirement if we want to keep control of the ontological layer, while keeping users free from modeling constraints. Filling this gap is the main task of a supervised content revision process. Nevertheless, to make the bottom-up approach plainly effective, users are encouraged to fit their lexical concepts and relations to the basic ontological choices and capture non-trivial aspects of their intended meanings.

The first enrichment procedure performed on the lexical resource was the association of each of 4,586 word senses (belonging to 1,111 fundamental noun lemmas having the highest rank in frequency lists of Italian language and covering about 90% of all textual occurrences) to a unique ontological category.

The experimentation was conducted by a group of graduate students of computational linguistics who worked separately on different sets of word senses and further discussed classification problems. The experimentation conducted using the Senso Comune platform was carried out in three phases:

(I) Unsupervised common sense classification;

(II) Revision of the classification (lead by Chiari, Vetere and Oltramari and four students) with the additional task of giving a confidence evaluation to the classification using three tags (accepted, controversial, not accepted) and discussion;

(III) Final revision of consistency in classification actions.

Since ontological categorization is not a simple task and involves complex metalinguistic and cognitive operations significant control check strategies were introduced by giving experimenters the possibility of associating a confidence labels to their choices asserting whether their classification was perceived as fully confident or problematic or ultimately very uncertain. We further checked inter-annotator agreement, and observed what categories and association tasks were mostly accepted as common by different annotators, what produced more disagreement, and what were perceived as hazardous.

One of the most interesting results regards the level of abstractness of the most problematic ontological categories that emerged in the annotation process. More specifically very abstract categories such as *entity, function, social object, state, idea* appear to be more controversial in the association process and generate more disagreement than categories that are located below in the hierarchy as *animal, person, artefact*. Detailed exposition of the results has been done in [11].

Discussion of results prompted the research group to allow multiple classifications of senses in the mapping annotation scheme, in order to evaluate specific patterns in possible associations, 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 emerged

during the experiment.

Figure 2. The annotation of confidence evaluation

Collaborative annotation was thus achieved by the inclusion of discussion strategies that aimed at increasing cooperativeness and awareness: (a) collective discussion, (b) explicit annotation of controversial cases and (c) confidence evaluation and with (d) man-machine tutoring tools (TMEO, §2.4).

2.4 Tutoring annotation tool: TMEO

Taking a collaborative perspective has played a key-role in phase I of the association experiment. In particular, we moved from some general considerations: experts are pretty effective in discriminating ontological categories and classify new concepts, but they embody just a tiny bit of the potential annotators available on the Web 2.0; conversely, training novice ontology annotators is time-consuming and engaging them in classification experiments can be tricky. A trade-off solution would be to build an interactive tool to support non-expert users and guarantee good-quality results in collaborative tasks of ontology classification. In Senso Comune, we've pursued this solution by using TMEO (Tutoring Methodology for the Enrichment of Ontologies) for the experiment with the undergraduate students. TMEO is a classification system based on broad foundational distinctions from DOLCE-Spray, a simplified version of DOLCE [12], which can be implemented with a synoptic map (Figure 3) or with a sequential question answering procedure (Figure 4): it helps the user/editor to select the most adequate category of the reference ontology as the super-class of the given lexicalised concept: different answer paths lead to different mappings between the lexicon and the (hidden) ontological layer.

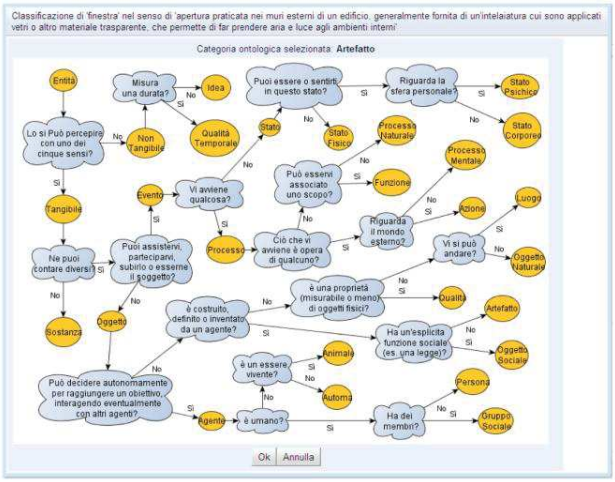

Figure 3. This conceptual map represents the Q/A mechanism underlying TMEO. DOLCE-Spray categories are represented in yellow circles; state transitions are driven by "yes/no" answers (black arrows) to questions enclosed in blue clouds.

After six months of experimentation with TMEO (including phase II revision), data were analysed to extract information about word sense distribution in ontological categories, classification problems, etc. A detailed discussion of the results is presented in [12]. In summary, 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 instability was raised by categories such as ENTITY, TANGIBLE, NON TANGIBLE, FUNCTION, OBJECT, STATE, IDEA [12]. By and large, the categorization of concrete entities was performed in a more accurate and confident way than the categorization of abstract notions. As a result of the experiment, we have 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.

2.5 Experimenting verbal frame integration

Figure 4. An example of the Q/A procedure in TMEO

Current work focuses of enriching the resource with data-induced verbal frames. The target corpus consists of about 8,000 usage examples associated with the fundamental senses of the verb lemmas in the resource.

The annotation task involves tagging the usage instances with syntactic and semantic information about the participants in the frame realized by the instances. Specifically, in syntactic annotation, annotators mark the beginning and end point of each constituent realized in the sentence, and tag them with information about the type of phrase and grammatical relation (3 broad dependency categories). In semantic annotation, users are asked to attach a semantic role and an ontological category to each participant and to annotate the sense definition associated with the filler when it is a noun. For this aim, we provide them with a hierarchical taxonomy of 24 coarse-grained semantic roles based on [13], together with definitions and examples for each role, as well as decision trees for the roles with rather subtler differences. As in the previous experiment of `ontologization' of noun senses (section 3.2), the TMEO methodology is used to help them selecting the ontological category in Senso Comune's top-level. For noun sense tagging, the annotator exploits the senses already available in the resource. Drawing on the results of the previous experiment on noun senses, we have decided to allow multiple classifications, that is, we allow the users to annotate more than one semantic role, ontological category and sense definition for each frame participant. Up to now about 400 usage examples (about 6% of the entire corpus) were each annotated 9 times by undergraduate students in a pilot experiment we performed to release the beta version of the annotation scheme. The interannotator agreement (Cohen's kappa) among those 9 annotators is 0.85 for the agreement on syntactic dependency category and 0.63 for the agreement on the semantic role (the ontological category and the noun sense tagging were not included in this pilot experiment). Such values are considered respectively as good and acceptable for notoriously difficult semantic tasks on skewed data [15]. Actually, for the semantic role, the data is so skewed that 5 rare roles do not appear at all in the experiment corpus, and among those that are present, the most frequent appears about 50 times more than the least frequent.

These good results having validated the annotation scheme, the annotation of the whole target corpus is currently going on. Fig. 5 illustrates the annotation of the semantic role of the subject of the target verb *leggere* in one usage example of its fourth sense ('amare la lettura, dedicarsi a essa').

Figure 5. The annotation of semantic roles

The basic idea behind the overall frame annotation effort is to use the annotated corpus as an empirical base to induce the corresponding verbal frames.
The resulting annotated corpus could be exploited as a source of (linguistic) knowledge in Natural Language Processing tasks such as semantic role labelling, word sense disambiguation, textual entailment and linguistic research on frame structures.

3. THE ROLE OF COLLABORATIVE APPROACHES IN SENSO COMUNE

As we have mentioned in previous sections, by separating the linguistic layer from the ontology, we allow language users to manifest their knowledge in a free, incremental, natural, collaborative and potentially conflicting way. As Wikipedia demonstrates, collaborative projects produce huge amount of knowledge, which is continuously updated, amended and extended by wiki-editors. We think that by applying a "crowdsourcing" approach to the collection of human common-sense and linguistic knowledge can also fit the "Semantic Web" paradigm. To build a semantic resource through a cooperative process, Senso Comune follows two main paths:

a) axiomatized top-level ontological categories and relations are introduced and maintained by ontologists in order to constrain the formal interpretation of lexicalised concepts (top-down direction);

b) users are asked to enrich the semantic resource with linguistic information through a collaborative approach (bottom-up direction). A special tutoring system supports users by interacting with users on the basis of question-answering mechanisms (i.e. TMEO, §2.4).

In order to exploit the collaborative potential of environment after the acquisition of the basic linguistic resource and its modeling, Senso Comune has been extended through a Web-based cooperative platform. The platform shares a number of key features with wikis:

- Editing through browser
- Rollback mechanism (versioning of saved changes is available, so that an incremental history of the same resource is maintained)
- Controlled access and different role read/write privileges
- Collaborative editing and discussion forum support
- Emphasis on linking
- Search functions

Senso Comune has also some relevant features that are significantly different from wikis:

- a rich interactive and WYSIWYG Web interface that is tailored to linguistic content;
- annotations are encoded directly into the Senso Comune meta-model;
- support of multiple annotation and conflict on annotation agreement signalling;
- support of annotation processes with the aid of the tutoring system supports on the basis of question-answering mechanisms – i.e. TMEO (A Tutoring Methodology for the Enrichment of Ontologies).

Senso Comune presently allows a number of annotation actions and user generated content: adding new lemmas, adding new word senses, associating word senses to multiple ontological categories, adding usage examples, annotating usage examples with syntactic and semantic information about the participants in the frame realized by the instances, including argument/adjunct distinction, adding and modifying lexical relations (synonymy, hyponymy, hyperonymy, meronymy, holonymy) to word senses.
As we have showed in §2.3, collaboration is also enhanced by the annotation of self-evaluative notes regarding the confidence annotators pose in the single tagging action performed, stimulating self-reflection and metalinguistic and metacognitive awareness and also discussion among different annotators.
Collaborative annotation is conceived within Senso Comune as a productive restructuring method for the overall architecture of the web-based environment, but also for the progressive refinement and reorganisation of the theoretical issues posed by the integration of the three main representation modules. Being annotation an interpretative task, it is a primary source for observing phenomena and re-planning links and relationships among layers.

Figure 6. The annotation of lexical relations

4. CONCLUSIONS

Interfacing ontologies with advanced linguistic technologies is the *conditio sine qua non* in order to allow effective machine understandability of "human meanings", and thus supporting non-trivial applications based on semantic technologies. By implementing a collaborative approach to linguistic knowledge acquisition, Senso Comune aims at providing open and high-quality resources to feed semantic applications. Moreover, collaborative construction of hybrid semantic resources may

support automated ontology learning, e.g. reducing incompleteness, inconsistency, ambiguity, and sparseness of the acquired ontologies. Relevant functions of collaborative procedures in Senso Comune platform include enabling faster and better annotation results; increasing the degree of critical discussion on annotating strategies and their driving ratio; providing significant feedback both to the development of better annotation methodologies and to the restructuring of the internal module that represents relationships and mapping among the different modules (top level, lexical, frame).

5. REFERENCES

[1] Fellbaum C. 1998. WordNet - An Electronic Lexical Database. MIT Press, Cambridge, Massachusetts.

[2] Assem M., Gangemi A, Schreiber G. 2006. Conversion of WordNet to a standard RDF/OWL representation. In *Proceedings of the International Conference on Language Resources and Evaluation* (LREC 2006) (Genoa, Italy, May 24-26, 2006).

[3] Pianta E., Bentivogli L., Girardi C. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the first International Conference on Global WordNet*. Central, Institute of Indian Languages, Mysore.

[4] Roventini A., Alonge A., Calzolari N., Magnini B., Bertagna F. 2000. ItalWordNet: a large semantic database for Italian. In *Proceedings of the International Conference on Language Resources and Evaluation* (LREC 2000) (Athens, Greece, May 31-June 2, 2000), 783–790.

[5] Lenci A., Johnson M., Lapesa G. 2010. Building an Italian framenet through semi-automatic corpus analysis. In *Proceedings of 7th International Conference on Language Resources and Evaluation* (LREC 2010) (La Valletta, Malta, May 17 -23, 2010).

[6] Tonelli S., Pighin D., Giuliano C., Pianta E. 2009. Semi-automatic development of FrameNet for Italian. In *Proceedings of the FrameNet workshop and masterclass*, Milano.

[7] De Mauro, T. 1980. Guida all'uso delle parole. Editori Riuniti, Roma.

[8] Oltramari, A. and Vetere, G. 2008. Lexicon and Ontology Interplay in Senso Comune. In *Proceedings of OntoLex 2008, 6th. International Conference on Language Resources and Evaluation* (LREC 2008) (Marrakech, Morocco, May 31, 2008).

[9] Oltramari, A., Vetere, G., Lenzerini, M, Gangemi, A., Guarino, N. 2010. Senso Comune. In *Proceedings of 7th International Conference on Language Resources and Evaluation* (LREC 2010) (La Valletta, Malta, May 17 -23, 2010).

[10] Suarez-Figueroa M.C., Gomez-Perez A., Motta E., Gangemi A. (eds). 2012. *Ontology engineering in a networked world*. Springer, Berlin.

[11] Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D. , Patel-Schneider, P. F. 2003. *The Description Logic Handbook: Theory, Implementation, Applications*. Cambridge University Press, Cambridge, UK,

[12] Chiari, I, Oltramari, A., and G. Vetere. 2012. Di che cosa parliamo quando parliamo fondamentale?, In *Atti del Convegno della Società di Linguistica Italiana* (Viterbo 27-29 settembre 2010) S. Ferreri, Ed. Roma: Bulzoni, 185-202.

[13] Bonial, C. and Corvey, W. and Palmer, M. and Petukhova, V.V. and Bunt, H. 2011. A Hierarchical Unification of LIRICS and VerbNet Semantic Roles. In *Proceedings of the 2011 IEEE Fifth International Conference on Semantic Computing*, 483-489.

[14] Oltramari A., Vetere G., Jezek E., Chiari I., Zanzotto F.M., Nissim M., Gangemi A 2013. Senso Comune: A Collaborative Knowledge Resource for Italian. In Gurevych, I. Kim, J. eds. *The People's Web Meets NLP: Collaboratively Constructed Language Resources*. Springer Verlag, 45-67.

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

[16] Nuzzolese A.G., Gangemi A., and Presutti V. 2011. Gathering Lexical Linked Data and Knowledge Patterns from FrameNet. O. Corcho, M. Musen, Eds. *Proceedings of K-CAP 2011, the Sixth International Conference on Knowledge Capture*, ACM.

[17] Gangemi A. 2010. What's in a Schema?, in *Ontology and the Lexicon*, Huang, C., Calzolari, N., Gangemi, A., Lenci, A., Oltramari, A., and Prevot, L., eds. Cambridge UP, 144-182.

[18] Baker, C.F., Fillmore, C.J., and Lowe, J.B. 1998. The Berkeley FrameNet project, in *Proceedings of the 1998 International Conference on Computational Linguistics* (COLING-ACL).

[19] Kipper, K., Dang, H.T., and Palmer, M. 2000. Class-Based Construction of a Verb Lexicon, In *AAAI-2000 Seventeenth National Conference on Artificial Intelligence*.

[20] Gadamer, H. G. 1960/2004. *Truth and Method*. 2nd rev. edition. Trans. J. Weinsheimer and D. G. Marshall. New York, Crossroad.