

Chapter 2

Senso Comune: A Collaborative Knowledge Resource for Italian

Alessandro Oltramari, Guido Vetere, Isabella Chiari, Elisabetta Jezek,
Fabio Massimo Zanzotto, Malvina Nissim, and Aldo Gangemi

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 download. A restricted community of contributors is currently working on increasing the lexical coverage of the resource.

A. Oltramari (✉)
Carnegie Mellon University, Pittsburgh, USA
e-mail: aoltrama@andrew.cmu.edu

G. Vetere
IBM Center for Advanced Studies of Rome, Rome, Italy
e-mail: gvetere@it.ibm.com

I. Chiari
Università *La Sapienza* di Roma, Roma, Italy
e-mail: isabella.chiari@uniroma1.it

E. Jezek
Università di Pavia, Pavia, Italy
e-mail: jezek@unipv.it

F.M. Zanzotto
Università di Roma *Tor Vergata*, Province of Rome, Italy
e-mail: zanzotto@info.uniroma2.it

M. Nissim
Università di Bologna, Bologna, Italy
e-mail: malvina.nissim@unibo.it

A. Gangemi
STLab – ISTC-CNR, Roma, Italy
e-mail: aldo.gangemi@cnr.it

2.1 Introduction

*Senso Comune*¹ is the project of building an open knowledge base for the Italian language, designed as a crowd-sourced initiative that stands on the solid ground of an ontological formalization and well-established lexical resources: in this respect, it leverages on Web 2.0 and Semantic Web technologies. The community behind this project is growing and the knowledge base is evolving by integrating collaboratively 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 download.

In the present contribution we provide an overview of the project, present some initial results, and discuss future directions. We firstly illustrate history and 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 model underlying the knowledge base, from the perspective of its ontological structure. This paper also focuses on the methodological aspects of the knowledge acquisition process, introducing an interactive Q/A system (TMEO) designed to help users assigning ontological categories to linguistic meanings. Finally, we report the results of the experiment on ontology tagging of noun senses in Sect. 2.5, and stress the relevance of the resource to Natural Language Processing 2.6.

2.1.1 History and Objectives

In fall 2006, a group of Italian researchers² 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 development of intelligent information systems capable of driving business and public services in Italy. Free, high quality lexical resources such as WordNet⁴ contribute to the growth of intelligent information systems in English speaking countries. Lexical machine-readable resources for Italian – primarily

¹www.sensocomune.it

²Besides the authors of this paper, the group, lead by Tullio De Mauro, includes Nicola Guarino, Maurizio Lenzerini and Laure Vieu.

³IBM Italia Foundation’s symposium *La dimensione semantica dell’Information Technology* (The semantic dimension of Information Technology), Rome, November 27, 2006.

⁴<http://wordnetweb.princeton.edu/perl/webwn>

MultiWordNet,⁵ EuroWordNet and the follow-up project SIMPLE⁶ – 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, as in the rest of the world, new communication technologies increase the pace of linguistic changes.

From the outset, *Senso Comune* was conceived as a linguistic knowledge base rather than a dictionary. It is actually based on a conceptual apparatus that is not usually present in standard linguistic resources. In particular, each sense is mapped to ontological categories, and is associated with semantic frames.

The starting point to build such a knowledge base has been the acquisition of a high-quality lexical resource, namely De Mauro’s ‘vocabolario di base’ (Basic Vocabulary), which consists in the 2,071 most frequent Italian words, kindly made available by the author. The Basic Vocabulary of Italian was developed by De Mauro in 1980 [11] and further updated with minor changes up to 2007. It contains three different vocabulary ranges, the first being the so called ‘fundamental vocabulary’ containing the top 2,000 lemmas with top rank in two frequency lists of Italian written (LIF) and spoken language (LIP) – see [5] and [12].

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. An interdisciplinary, cross-organization team hosted at the Center for Advanced Studies of IBM Italia started designing a representational model 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 2009, after 1 year of work.

Based on the acquired resource (see Sect. 2.3), 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. That work was carried out by undergraduate students under the supervision of the association researchers (see Sect. 2.5).

The development of *Senso Comune* has followed two main tracks so far. 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

⁵See for example: <http://multiwordnet.itc.it/english/home>

⁶<http://www.ub.edu/gilcub/SIMPLE/simple.html>

studying how to extend the model to encompass the kind of lexical knowledge that 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 [48].

2.1.2 *General Linguistic Perspective*

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 expressions, 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 tenet is that natural languages manifest themselves in actual usage scenarios, while the regularities that those languages show are a consequence of social evolution and consensus. Since languages serve humans in dealing with the world, ontologies (i.e., theories about physical, social or abstract realities) constitute a reference to characterize social evolution and consensus of language with respect to extra-linguistic entities. In other words, although language is far from being a mere “picture of reality”, theories about 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 [33]. In our first experiment we chose to mark linguistic data with concepts of a general formal ontology.

Ontologies represent an important bridge between knowledge representation and computational lexical semantics, and form a continuum with semantic lexicons [20]. The most relevant areas of interest in this context are Semantic Web and Human-Language Technologies: they converge in the task of pinpointing knowledge contents, although focusing on two different dimensions, i.e. 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 straightforwardly related to ontological categories and relations. Polysemy refers to a genuine lexical phenomenon that is

generally absent in well-formed ontologies; the formal features of computational lexicons are far from being easily encoded in a logic-based language.⁷

Since the early 1980s, there has 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. [31] 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.⁸ In the following section we quickly survey three of the most important state-of-the-art computational lexicons, i.e. WordNet, FrameNet and VerbNet, providing the general conceptual framework in which *Senso Comune* is rooted.

2.1.3 Comparing *Senso Comune* with WordNet, FrameNet, and VerbNet

WordNet was developed in Princeton University under the direction of the famous cognitive psychologist 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” ([13], p.7). WordNet is constituted by synsets (lexical concepts), namely set of synonym terms – e.g. (life form, organism, being, living thing). The idea of representing world knowledge through a semantic network (whose nodes are synsets, and whose arcs are lexical semantic relations⁹) 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 in the most recent releases), and substantial improvements of the entire WordNet architecture, aimed at facilitating hierarchical organization and computational tractability. Accordingly, RDF- and OWL-based implementations have been released (e.g. [1]).

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 fruitful attempts to annotate WordNet with domain/topical information in order to improve the overall accessibility to the dense lexical database. Wordnets have been and are being constructed in dozens of

⁷Lexicons often omit any reference to ontological categories that are not lexicalised in a language, although it sometimes happens, as with EuroWordNet’s ILIs or FrameNet’s non-lexicalised frames.

⁸In this respect, our approach is essentially different from OntoNotes [32], 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” [33] while *Senso Comune*, conversely, is explicitly grounded on an ontological model (see Sect. 2.2).

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

languages. Besides the EuroWordNet project that built wordnets for eight European languages, BalkaNet project,¹⁰ encompassing six languages, and PersiaNet,¹¹ have been developed. In addition, wordnets are being constructed in Asia and South America.¹² It's also worthwhile to mention the SIMPLE project [19], an evolution of the EuroWordNet project, which implements Pustejovsky's qualia roles [34].

WordNet has been often considered as a lexical ontology or at least as containing ontological information: although synsets can be conceived as lexically grounded counterparts of ontological categories, wordnet-like resources do not rely on any explicit logical infrastructure.

Senso Comune has borrowed from WordNet many basic intuitions about lexical ontology. However, *Senso Comune* differs from WordNet in many respects. Firstly, besides focusing on synonymy and hyponymy relations with the aim of bringing out the conceptual structure behind the lexicon, *Senso Comune* also adopts a set of a priori ontological distinctions, to identify the ontological commitments behind each sense. Secondly, *Senso Comune* will also contain a parallel structuring based on frames. A semantic lexicon can be structured from a different perspective, focusing on semantic frames instead of synsets, as in the case of FrameNet [39]. 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 cognitively-grounded constructs carrying several kinds of information: the structure of the frame itself, how to use the frame, what one can expect to happen after the occurrence of that frame, and what to do if these expectations are not confirmed [25]. There is a close kinship between AI or cognitive frames and linguistic-based semantic frames: a comprehensive analysis of their relations is presented in [15].

FrameNet is the most comprehensive repository of semantic frames; it 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 [14], FrameNet aims at documenting "the range of semantic and syntactic combinatorial possibilities (valences) of each word in each of its senses" through corpus-based annotation. For example, the **Discussion** frame, namely an abstraction of situations where discussants talk about something in a given place at a given time, is grounded in several lexical occurrences in the FrameNet corpus, which are lemmatized as "lexemes", which are grouped into "lexical units" – LUs: e.g. the noun *negotiation* or the verb *debate*. A frame also has different semantic roles (or "frame elements" – FEs): e.g. *Interlocutor* or *Topic*. On their turn, semantic roles are grounded, e.g. the nouns *president* and *advisor* ground the *Interlocutor* role in the **Discussion** frame. The same LU may ground distinct frames or semantic roles: the noun *president*, for example, also grounds the **People** frame.

FrameNet contains about 12,000 LUs in about 1,000 frames (grounded in lexemes from about 150,000 annotated sentences). As with WordNet, new projects

¹⁰<http://www.ceid.upatras.gr/Balkanet/>

¹¹<http://persianet.us/>

¹²For an updated list of wordnet projects see: <http://www.globalwordnet.org/>

are under development to yield FrameNet-based computational lexicons for other languages: SALSA project in Germany,¹³ Japanese FrameNet,¹⁴ and domain specific resources like the Soccer FrameNet.¹⁵ FrameNet has also been ported to RDF-OWL, and aligned to WordNet for interoperability [26].

Senso Comune's model is being extended to encompass verbal frames (see below (see below and Sect. 2.6), which will make it comparable to existing framenet-like resources. However, existing framenets don't supply a formal characterization of the relations between frames, roles, etc., although FrameNet documentation is more explicit than WordNet's about its possible formal interpretation. In practice, such interpretation has to be reconstructed (cf. [26]). On the contrary, formal interpretation of lexical knowledge is a key feature of *Senso Comune*.

FrameNet is not the only resource for semantic frames and roles we are reusing for building the frame-oriented structuring of *Senso Comune*. VerbNet [18] is a freely available verb lexicon which encodes syntactic and semantic information for *classes* of verbs, and is linked to WordNet and FrameNet. Verb classes are mainly based on Levin's classification [22], thus implying a strong link between the syntax and the semantics of verbs. Indeed, in VerbNet, the semantics of a verb is associated with its syntactic frames, and information about thematic roles and selectional preferences is also included. Verbs belonging to the same VerbNet class are supposed to share the same subcategorisation frame – information that is not included in FrameNet – and have the same selectional preferences and thematic roles associated with the expected arguments.

While there are a few Italian wordnets available (e.g. MultiWordNet [30] and ItalWordNet [38]),¹⁶ and there have been attempts at automatically inducing an Italian FrameNet [21, 47], there is as yet no VerbNet-like resource for Italian. However, as a starting point, *Senso Comune*'s predicate representation has been based on efforts towards combining theoretical and corpus-derived information for obtaining a verb classification which is meaningful at the syntax-semantics interface: in particular, [35] combines a theoretical approach grounded on Pustejovsky's Generative Lexicon [34] and a corpus-based distributional analysis for representing word meaning.

2.2 The Model

The adoption of a full-featured, legacy dictionary as a foundation for the resource construction, has led to modeling *Senso Comune* basing on a clear distinction between lexicographic structures and linguistic facts. Basically, *Senso Comune*'s

¹³<http://www.coli.uni-saarland.de/projects/salsa>

¹⁴<http://jfn.st.hc.keio.ac.jp/>

¹⁵<http://www.kicktionary.de/>

¹⁶There is yet another WordNet for Italian developed by a company, but it is not freely accessible.

notion of LEMMA captures the section of a dictionary where an etymologically consistent bundle of senses (that we call MEANING RECORD) of a given lexeme is described by means of a suitable lexicographic apparatus (e.g. definition, grammatic constraints, usage examples). Thus, although related, it must not be confused with the linguistic notion of *lexeme*. This is a distinguishing feature of *Senso Comune* with respect to other models, such as LMF [6] or Lemon [8], to which, however, *Senso Comune* is strongly connected. The common goal of these models is to provide a structure to accommodate semasiological information, i.e. linguistic resources where lexical units are associated with their acceptations. Separating the description of linguistic senses and relationships (e.g. synonymy, hyponymy, and antinomy) from the formal account of their phenomenal counterparts (e.g. concepts, equivalence, inclusion, disjointness) brings a number of benefits. Primarily, this separation prevents lexicographical artifacts to be directly mapped to logic propositions, thus relieves the dictionary the burden of embodying ontological commitments [48], while preserving the possibility of relating lexicographic records with any suitable ontology.

Senso Comune's model is specified in a set of "networked" ontologies [45] 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 modeling the predicative structure of verbs and nouns. The root of the class hierarchy of *Senso Comune* is ENTITY, which is defined as the class of anything that is identifiable by humans as an object of experience or thought. The first distinction is among CONCRETE ENTITY, i.e. the class of objects located in definite spatial regions, and NON PHYSICAL ENTITY, including objects that don't have proper spatial properties. In the line of [43], CONCRETE ENTITY is further distinguished into 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. The *top level* ontology is inspired by DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [24], which has been developed in order to address core cognitive and linguistic features of common sense knowledge. We kept the basic ontological distinctions: DOLCE's *Endurant* and *Perdurant* match *Senso Comune*'s CONTINUANT and OCCURRENT, respectively. The main difference between *Senso Comune*'s top level and DOLCE is the merging of DOLCE's *Abstract* (e.g. mathematical entities, dimensional regions, ideas) and *Non-physical-endurant* (e.g. social objects) categories into a *Senso Comune* category NON PHYSICAL ENTITY.

Among non physical entities, *Senso Comune*'s top level distinguishes CHARACTERIZATION, which is defined on the basis of the irreflexive, antisymmetric relation CHARACTERIZES, that maps instances of non physical entities to other entities (including collective ones), meaning that the former represent some aspect of the latter in some way and under some respect. SOCIAL OBJECT is the class of non physical entities instituted within (and dependent upon) human societies e.g. by means of linguistic acts [40], while INFORMATION OBJECT is the class of social objects which convey information of any kind.

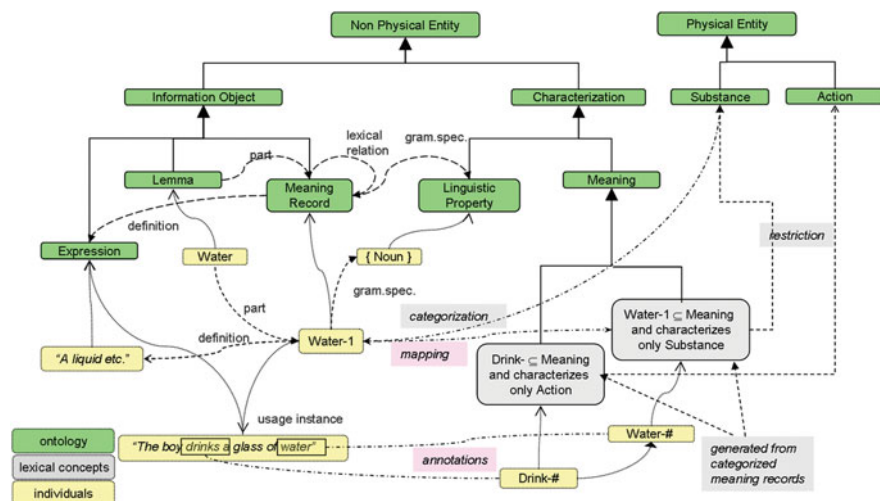


Fig. 2.1 *Senso Comune* model

The semasiological model of *Senso Comune* (Fig. 2.1) unfolds under the hierarchy of non physical entities. In particular, LEMMA and MEANING RECORD are both information objects, the latter part of the former, whose instances, along with their attributes, form the main body of our lexical resource. On the other hand, MEANING is a social characterization, whose instances occur in the context of linguistic acts. A specific meaning (e.g. *water* in the sense of *liquid substance*) will be a subclass of MEANING, suitably restricted to characterize only liquid substances. The instance of MEANING RECORD where such meaning is described, will be mapped to that class. Mapping between instances of meaning record and meaning classes can be done, in the OWL2 syntax, by annotations, punning, or other structures. In any case, formal semantics of mappings can be specified in different ways, which are out of the scope of this writing. Attributes of meaning record instances (e.g. glosses, grammatic features, usage marks, rhetoric marks, etymology, etc) do not affect the mapped meaning class (if any). Moreover, different meaning records instances (e.g. from different dictionaries) can be mapped to the same meaning class. This way, the model may accommodate meaning records coming from different sources, that might use different sets of attributes (e.g. different usage marks). Also, lexical relations are predicated on meaning records (instead of meanings); hence they are set among information objects and do not have a direct ontological import. Any correspondence (e.g. hyponymy \mapsto inclusion) should be introduced based on suitable heuristics. In sum, both meaning and lexical relation records are purely informative, which could facilitate the process of integrating different (possibly diverging) sources of lexical knowledge.

By separating linguistic from formal semantic features, *Senso Comune* allows users to express their knowledge in a free and natural way. This implies, however, the potential rise of conflicts and disagreements. For instance, synonymy or polysemy of words can be perceived differently by different users. Platforms like Wikipedia provide means for amending errors and arbitrating conflicts, based on self-regulation emerging from large (and presumably well behaved) user communities. We think that a collaborative approach can be also adopted when collecting linguistic and semantic knowledge. At the same time, we recognize that such knowledge requires a specific treatment. On the one hand, linguistic knowledge is less sensitive to emotive opinion clashes or prejudice than encyclopedic one (e.g. about people or facts); on the other hand, in order to take the maximal advantage from user input, we need a formal apparatus that works behind the curtains.

To build a semantic resource through a cooperative process, *Senso Comune* follows two main paths:

- **Top-down** axiomatized top-level ontological categories and relations are introduced and maintained by ontologists in order to constrain the formal interpretation of lexicalised concepts;
- **Bottom-up** language users are asked to enrich the semantic resource with linguistic information through a collaborative approach.

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.

For this reason, we designed TMEO, a tutoring methodology to support enrichment of hybrid semantic resources based on *Senso Comune*'s ontological distinctions (see Sect. 2.4). In the rest of this paper we present some aspects related to the population of the *Senso Comune*'s knowledge base, focusing both on the top-down and the bottom-up approach (Sects. 2.3 and 2.4, respectively).

2.3 The Acquisition Process

Senso Comune's knowledge base has been populated with approximately 13,000 meaning entries (senses) generated by acceptations of 2,075 lemmas from the De Mauro's core Italian dictionary [10]. Starting from this set of fundamental

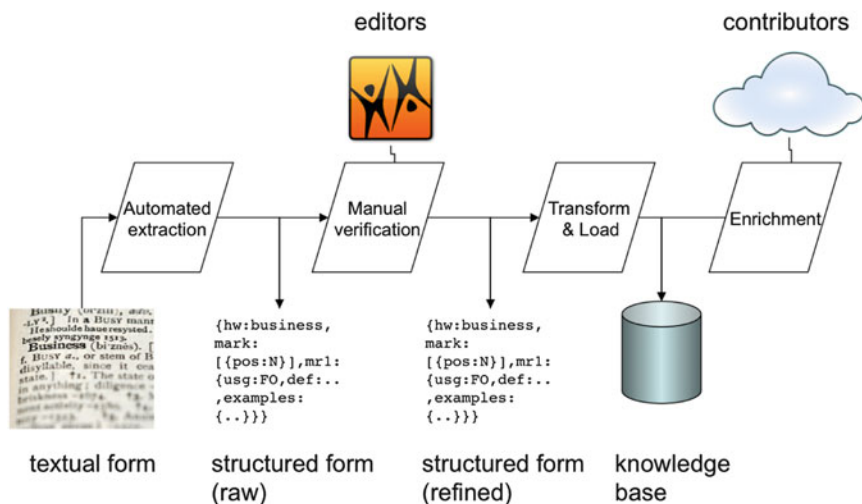


Fig. 2.2 Acquiring the basic lexicon

senses, the *Senso Comune* knowledge base is developed by the supervised contribution of speakers through a cooperative open platform.

2.3.1 Acquiring the Basic Lexicon

Starting from plain textual lemmas extracted from De Mauro's dictionary,¹⁷ the acquisition process of *Senso Comune* consisted in producing individuals corresponding to some of the main classes of the *Senso Comune*'s lexical ontology: **LexicalEntry**, **Word**, **MeaningRecord**, and **UsageInstance** classes. This conversion turned out to be less trivial than initially expected, since lexicographers are used to use the same typographic conventions to convey information that is assigned to different portions of the *Senso Comune* target model. For example, senses and usage instances are not always clearly distinguishable, especially in presence of several meaning 'nuances', which is quite common for basic lemmas (Fig. 2.2).

Therefore, after having automatically transformed the dictionary content into an intermediate XML format, a manual revision was needed to amend errors. In many cases, corrections required significant linguistic skills.¹⁸

¹⁷Grande dizionario italiano dell'uso (Gradit), Torino, UTET

¹⁸Two teams of five linguists each, based in Rome and Bologna, under the supervision of Isabella Chiari and Malvina Nissim, were dedicated to this task.

2.3.2 The Cooperative Platform

After the acquisition of the basic terminology, *Senso Comune* has been extended through a Web-based cooperative platform. The platform shares a number of key features with wikis:

- Editing through browser: contents are usually inserted through web-browsers with no need of specific software plug-ins.
- Rollback mechanism: versioning of saved changes is available, so that an incremental history of the same resource is maintained.
- Controlled access: even if, in most cases, wikis are free access resources and visitors have the same editing privileges, specific resources (or parts of them) can be somehow preserved.
- Collaborative editing: many wiki systems provide support for editing through discussion forums, change indexes, etc.
- Emphasis on *linking*: resources are usually strongly connected to one another.
- Search functions: rich search functionalities over internal contents.

At the same time, *Senso Comune* shares some critical aspect with wikis:

1. Quality of contents: this aspect focuses on ‘bad’ or low-level contents.
2. Exposure to “malevolent attacks” that aim at damaging contents or at introducing offensive (or out of scope) information.
3. Neutrality: the difficulty of being completely fair when making statements about questionable matters. Even if linguistic meanings are less sensitive to neutrality than generic wiki contents, moderators are in charge of monitoring contents and behaviors.

With respect to Wiktionary,¹⁹ the Wikimedia project aiming at building open multilingual dictionaries with meanings, etymologies, pronunciations, etc., *Senso Comune* has the following differentiating features:

- Model: Wiktionary encodes each lemma in a wiki page, where different senses are coded as free text without specific identifiers. This choice makes hard to recover the conceptual information associated with lemmas. On the contrary, senses (and their relationships) are first-class citizens in *Senso Comune*.
- Interface: while Wiktionary is based on a generic wiki environment, *Senso Comune* has developed a rich interactive and WYSIWYG Web interface that is tailored to linguistic content (see Fig. 2.3).

Use cases of *Senso Comune*, however, are very close to Wiktionary’s ones. After searching a word, and visualizing the information obtained from the platform, users can decide whether to insert a new lemma, a new sense, a new lexical relation, or simply to leave a “feedback” (e.g. their familiarity with available senses and lexical relations). On the contrary, the deep conceptual part of the lexicon (the ontology) is

¹⁹<http://www.wiktionary.org/>

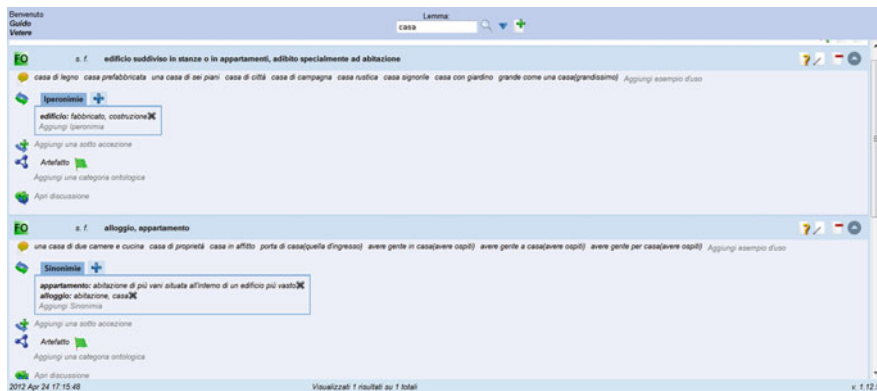


Fig. 2.3 The interface of *Senso Comune*

not accessible to users: when a new sense of a lemma is added, the system creates a corresponding specific concept to be positioned with respect to the ontological layer of the knowledge base. Then, possibly with the help of TMEO (see Sect. 2.4), the user can assign an ontological classification to the new sense. The current prototype of the *Senso Comune* computational lexicon is based on a relational database resulting from the linguistic model (see Sect. 2.2). The database has been also integrated with a DL-Lite reasoner [2], designed and implemented to operate on large ontologies.

2.4 The TMEO Methodology

In this section we introduce the general features of TMEO [27], a tutoring methodology to support semi-automatic ontology learning by means of interactive enrichment of ontologies (both from the lexical and the ontological levels).

TMEO is based on the simplified version of DOLCE adopted by *Senso Comune* (see Sect. 2.2). TMEO is inspired by Plato's dialectic (Socratic methodology to drive his disciples to true knowledge, posing questions and arguing on answers [36]): it exploits some suitable ontological properties for posing questions to users in support of domain independent or dependent knowledge modeling. TMEO is an interactive Q/A system based on general distinctions embedded in *Senso Comune*'s ontology.

Consider the case in which a given user is asked to classify the term (shoe), whose WordNet gloss is “footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material”.²⁰ After initializing TMEO wizard, the HCI interface will put the user through a series of intuitive conceptual questions – driven by the underlying ontological model – in

²⁰See <http://wordnetweb.princeton.edu/perl/webwn>

order to make explicit the intended meaning of the term. The following sequence reflects an experimental trial made with multiple users.

- TMEO-Wizard: Can you touch or see or smell or taste or hear or feel **a shoe**?
User: Yes
- TMEO-Wizard: Can you count or enumerate **shoes**?
User: Yes
- TMEO-Wizard: Can you say that “a **shoe** is happening or occurring”?
User: No
- TMEO-Wizard: Can you say that a **shoe** is a product of human activity (no matter whether hand-made or with the help of suitable tools and machines)?
User: Yes
- TMEO-Wizard: **shoe** in the sense of “footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material” has been classified as a kind of ARTIFACT.
- TMEO-Wizard: What are **shoes** commonly used for?
User: to walk
- TMEO-Wizard: Select the correct sense of **to walk**: [$s_1 - s_2 - s_3 - s_n$].²¹
User: s_1
- TMEO-Wizard: **to walk** in the sense s_1 has been classified as the typical FUNCTION of **shoes**.

Here the algorithm drives the user through tracing the following path of knowledge: *shoes as ARTIFACT have the common FUNCTION of being used in walking events*. As the above-mentioned scenario suggests, TMEO methodology may therefore be adopted not only in the unilateral classification of a given term (‘shoe’) but also in making related lexical items explicit. This kind of relatedness between terms actually unwraps the inter-categorical relation(s) holding between the corresponding ontological categories. Indeed, from the ontological viewpoint we can say that there is a relation of *Participation* holding between the category ARTIFACT (which is a kind of PHYSICAL OBJECT) and FUNCTION, which is conceptualized in *Senso Comune* as a kind of PROCESS.²²

TMEO has been implemented as a finite state machine (FSM): in general, the elaboration process of a FSM begins from one of the states (called a ‘start state’), goes through transitions depending on input to different states and can end in any of those available (only the subset of so-called ‘accept states’ mark a successful flow of operation). In the architectural framework of TMEO, the ‘start state’ is equivalent to the top-most category ENTITY, the ‘transitional states’ correspond to disjunctions within ontological categories and ‘accept states’ are played by the most specific categories of the model, i.e. ‘leaves’ of the relative taxonomical structure.

²¹For the sake of readability, we don’t go through the basic senses of the verb ‘to walk’, also assuming that s_1 is adequately selected by the user.

²²Note that we may wish to distinguish descriptions of functions from actual ones, namely those functions which are performed at a certain time by a given object. In the above example we simplify this distinction only focusing on the latter case.

Classificazione di 'cane' nel senso di 'animale domestico molto comune, diffuso in tutto il mondo, usato per la caccia, la difesa, nella pastorizia, come animale da compagnia o per altre attività'

Entità

Puoi percepire cane con almeno uno dei cinque sensi?

☐ Y

☐ N

Avanti Annulla

Non classificato

Aggiungi una categoria ontologica

Apri discussione

s. m. sing. in espressioni negative, nessuno

2011 Jan 29 16:32:27 Visualizzati 1 risultati su 1 totali v. 1.9.0

Senso Comune contiene attualmente:

- 31945 lemmi; di cui 2059 con almeno un'accezione
- 14046 accezioni; di cui 13161 fondamentali 7 di alto uso 73 di alta disponibilità 188 comuni 412 tecnico specialistiche
- 586 relazioni lessicali

Fig. 2.4 Senso Comune’s interface for TMEO-Wizard. Users can classify word-senses by answering to a logically-interconnected sequence of questions, designed on *top* of *Senso Comune* ontology

In this context, queries represent the conceptual means to transition: this means that, when the user answers to questions like the ones presented in the above-mentioned example (e.g. “can you count or enumerate shoes?”), the FSM shifts from one state to another according to answers driven by boolean logic²³. If no more questions are posited to the user, then this implies that the operations have reached one of the available final ‘accept state’, corresponding to the level where ontological categories don’t have further specializations (no transitions are left).

TMEO human language interface is very simple and comes in the form of a window where *yes/no* options are presented together with the step-by-step questions: Fig. 2.4 shows an example in Italian for the word ‘cane’ (=dog), where the Wizard asks whether one can perceive *cane* with the five senses or not. At the end of any single process of enrichment, the system automatically stores the new concept as an OWL class in the knowledge base under the ontological category selected by the user (e.g. in this sense, ‘shoe’ and ‘dog’ become respectively a subclass of ARTIFACT and of ANIMAL).

Future work on TMEO aims at extending the coverage of the model, adding new ‘transitional states’ and ‘accept states’. We discovered that users, in fact, have a high degree of confidence and precision in classifying the concepts referring to the physical realm, while they face several problems in distinguishing abstract notions like ‘number’, ‘thought’, ‘beauty’, ‘duration’, etc. (see Sect. 2.5): future releases of TMEO need to be improved both conceptually and heuristically, in this direction.

²³Uncertainty will be included only in future releases of the TMEO system.

2.5 Experiments on Noun Word Sense Ontology Tagging

An experiment on the association of word senses and ontological categories has been carried out using both a common sense direct tagging, and the TMEO tutoring tool in order to test advantages and disadvantages of bottom-up population of *Senso Comune*. The experiment aimed at observing procedures of association of word senses with 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 having the highest rank in frequency lists of Italian language and covering about 80 % of all textual occurrences) to a unique ontological category.

The work was carried out by a group of graduate students of Isabella Chiari's computational linguistics class at University of Rome La Sapienza. The procedure was carried out in three phases: (I) Primary unsupervised common sense classification lead by 12 students; (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.

For the annotation of ontological categories, experienced users directly select a single item from a given list containing all ontological categories. Categories can be also kept “opaque” in order to facilitate those who need guidance in understanding ontological commitments behind specific categorization choices. Thus students who were not confident in direct selection were advised to rely on TMEO. 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 lexicalised concept: different answer paths lead to different mappings between the lexicon and the (hidden) ontological layer (Fig. 2.5).

Since ontological categorization is not a simple task and involves complex metalinguistic and cognitive operations a significant control check was introduced by giving experimenters the possibility of associating a confidence label to their choices asserting whether their classification was perceived as fully confident or problematic – especially if the subject was in doubt among different possible categories – or ultimately tentative. We further checked inter-annotator agreement, and observed what categories and association tasks were accepted as common by different annotators, what produced disagreement, and what were perceived as hazardous. Contradictions and disagreements can emerge at the level of language – as stressed in Sect. 2.2 – and even more so in the task of ontological classification. Accordingly, we allowed the users to access to a dedicated ‘Forum’ room where they could discuss their ontological classification tasks, share their opinions and choices, ask moderators for advice if needed. In general, the Forum became the core tool of support for the experiment and a good instrument to monitor the learning progress of the subjects.

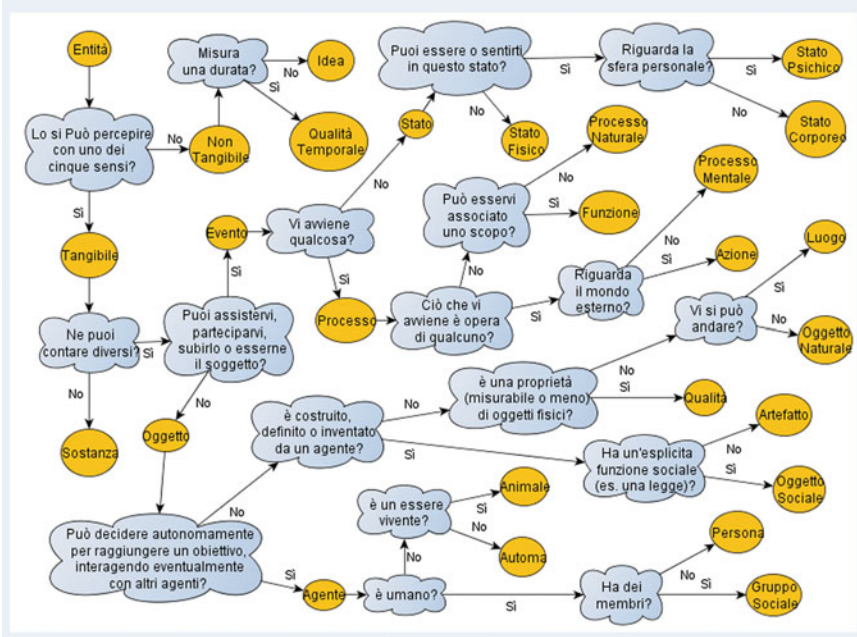


Fig. 2.5 This conceptual map represents the Q/A mechanism underlying TMEO. *Senso Comune* categories are represented in *yellow circles* with the corresponding Italian labels (literally translated from English, except for ‘Tangibile’ and ‘Non-Tangibile’ that maps respectively, to CONCRETE-ENTITY and ABSTRACT-ENTITY). State transitions are driven by “yes-no” answers (black arrows) to questions enclosed in blue clouds

After 6 months of work, including supervision, data was analysed 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.

Table 2.1 shows the most populated ontological categories, and the number of word senses attributed to them. The interpretation of this table is very complex and involves the consideration of the hierarchical structure of ontological categories and the observed preference for association of a basic (medium abstraction) level exhibited by the experimenters.²⁴

Further issues to be considered carefully are posed by the different degrees of confidence in the association process performed as well as inter-annotator agreement issues: 2,685 (59%) dictionary word senses were classified with full confidence, while 1,537 (33%) caused discussions, uncertainty and disagreement among annotators, and 364 (8%) revealed the ontology to be incomplete or

²⁴Please refer to [9] for a more complete version of Table 2.1 and a more detailed analysis of this experiment.

Table 2.1 Word senses attribution to ontological categories

Ontological category	WS	%
IDEA	689	15.02
ARTIFACT	505	11.01
PERSON	502	10.95
QUALITY	433	9.44
ACTION	413	9.01
NATURAL OBJECT	205	4.47
PSYCHOLOGICAL STATE	185	4.03
TEMPORAL QUALITY	184	4.01
EVENT	172	3.75

problematic. A confidence index and the evaluation of inter-annotator agreement are capital steps in the interpretation of tagging of all sorts performed by non-specialists giving an invaluable insight into complex cognitive and (meta)linguistic processes.

The data we collected shows that some ontological categories posed more association issues than others (from 68 to 81 %). For example, while ANIMAL, PERSON, NATURAL OBJECT, ARTIFACT, SUBSTANCE, and ACTION did not pose many confidence issues, a high degree of discussion and classification instability was raised by categories such as ENTITY, CONCRETE-ENTITY, ABSTRACT-ENTITY, FUNCTION, OBJECT, STATE, IDEA, which are mostly abstract categories. Further results lead us to observe the complex relationship among word senses as coded in traditional lexical resources as the dictionary used in the experiment and ontological categories: the richness or variety of ontological classes associated with each lemma entry. We have observed that there is a proportional relation between the number of word senses of a lemma and the variety of ontological categories. Most lemmas were associated to two or three different ontological categories while bearing an average of three to five word senses. Lemmas associated to only one ontological category in all the word senses are only 182 (20 % of all fundamental nouns), mostly belonging to PERSON (52), ARTIFACT (27), IDEA (18) and ACTION (14), like in the Italian lemmas *balcone* “balcony”, *calza* “socks”, *coltello* “knife”, *ingegnere* “engineer”, etc.

As a result of the experiment, the research group decided to allow multiple classifications of senses in further experiments, 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.

2.6 Relevance to Natural Language Processing

Resources such as WordNet, FrameNet and VerbNet are in constant development so as to increase their coverage and optimise their internal coherence. These efforts are more than welcome and encouraged within the Natural Language Processing

(NLP) community since such resources constitute a crucial supply of knowledge to be integrated in NLP systems. For instance, automatic word sense disambiguation (WSD) systems, and thus all the higher level NLP tasks that need WSD as a component, heavily rely on WordNet-like resources for creating gold standards and for system development. WordNet has also proved useful, for example in learning information extraction patterns for data mining [44], estimating semantic relatedness of concepts [29], and clustering entities for predicting violations of selectional restrictions [37]. In the latter respect, though, recent work has shown that learning selectional preferences from data using a distributionally-based algorithm can perform better than relying on hand-crafted resources such as WordNet [28].

Another specific NLP task that has hugely benefited from the resources we are discussing, and FrameNet in particular, is semantic role labelling (SRL), i.e. the identification and labelling of predicate arguments in text in an automated way. After the pioneering work of Gildea and Jurafsky [16], who indeed use FrameNet for training their SRL system, several shared tasks have been organised (for an overview see [23]). Interestingly, this task has also been tackled by *combining* WordNet, VerbNet, and FrameNet so as to make up for the shortcomings of each resource since they are complementary in the information they provide [17]. Shi and Mihalcea [42] indeed combine the three resources in order to enhance each of them and show, as a case in point, that they can perform robust semantic parsing this way.

WordNet, VerbNet, and FrameNet have undoubtedly proved a useful source of knowledge for NLP tasks. However, their main drawback is the fact that they are handcrafted, thus requiring a huge amount of manual work and resources, in time and economic terms. But we can look at the other side of the medal: while such resources are crucial for the development of semantically-aware NLP systems, it is also true that NLP tools can be used for *building*, or enhancing, such resources, especially in a semi-automatic, human-assisted setting, thus reducing the amount of human intervention. Inducing FrameNet-like structures has been the successful focus of large-scale projects like SALSA [7], for German, and we have already mentioned the existing efforts for inducing an Italian FrameNet [21, 47]. Work on Italian has also prompted an infrastructure for extending FrameNet induction to other languages [46].

Senso Comune lies on both sides of the medal as it will provide a lexical resource along with an annotated corpus associated with it that is used to improve the resource. In line with the rest of the activity, the linguistic annotation on the corpus is done with crowdsourcing methods (cf. Sect. 2.3.2). 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, including argument/adjunct distinction. Specifically, syntactic annotation involves identifying the constituents that hold a relation with the target verb, classifying them as arguments or adjuncts and tagging them with information about the type of phrase and grammatical relation. 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. For this aim, we provide

them with a hierarchical taxonomy of 27 coarse-grained semantic roles based on [4], 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 (Sect. 1.5), the TMEO methodology is used to help them selecting the ontological category in *Senso Comune*’s top-level (Sect. 1.4). 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 allow multiple classification, 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 we annotated about 400 usage examples (about 6 % of the entire corpus) in a pilot experiment we performed to release the beta version of the annotation scheme.

It is interesting to note that in spite of the difficulties related to specialised annotation, such as specific linguistic phenomena, current efforts towards using crowdsourcing methods for gathering linguistic annotation are proving successful (e.g. [3]), although the most technical information is usually added by experts. Also, thanks to regularly increasing amounts of annotated data, NLP tools can be used for inducing some of the annotation, possibly using active learning techniques, successfully employed in minimizing the annotation effort while maximizing accuracy and coverage for several NLP tasks [41]. This bootstrapping setting is already on the other side of the medal, since the resource is being used for developing semantically-aware NLP systems.

By being built collaboratively on the basis of a logically and linguistically motivated paradigm, and by being made freely available to the research community, *Senso Comune* can contribute to the virtuous cycle of using annotating data for developing and/or enhancing NLP systems and viceversa.

Moreover, by integrating in one resource several levels of representation, it encompasses the kind of information provided by the three different resources WordNet, VerbNet, and FrameNet.

2.7 Conclusions and Future Work

This paper presented *Senso Comune*, an open cooperative platform for the Italian language aimed at knowledge acquisition, and we discussed some of the major topics related to linguistic knowledge acquisition.

One of the main features of *Senso Comune* is the semiotic approach used to interface linguistic meanings and ontological concepts. Meanings are not modeled as concepts, but rather as signs. Accordingly, lexical relationships such as synonymy or hyponymy are not mapped into formal relations such as equivalence or inclusion, but are taken as input for the construction of ontological theories.

Thanks to the loose relation between linguistic and ontological data, conflicts and inconsistencies in user inputs do not affect the ontology directly; instead, there’s

room for introducing automatic, semi-automatic, or manual procedures to map linguistic senses to their ontological counterparts.

Current research includes modeling situations by means of frame-like structures, consistently with the formal model that is being developed. Lexical relationships to capture thematic roles will be therefore introduced. Another research direction is toward algorithms for automating the introduction of ontology axioms (e.g. equivalence, inclusion, disjointness, participation) based on linguistic information, by taking both quantitative and qualitative aspects into account.

Hybridisation of manual and crowd-sourced techniques for lexical knowledge acquisition, together with the contribution of NLP methods is also under study. Future efforts will be also devoted to widen the scope of the project, e.g. porting *Senso Comune* into the ‘Multilingual Semantic Web’ framework,²⁵ in order to enable cross-linguistic access and queries thorough Linked Data representations.

Finally, we think that *Senso Comune* as an open source of knowledge of Italian language can make a long way as key enabling factor for business, Web communities, and public services in Italy. The resource will be distributed under Creative Commons license and made available for any kind of use.

References

1. Assem Mv, 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), Genoa, Italy
2. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (eds) (2003) The description logic handbook: theory, implementation and applications. Cambridge University Press, New York
3. Basile V, Bos J, Evang K, Venhuizen N (2012) Developing a large semantically annotated corpus. In: To appear in proceedings of LREC, Istanbul, Turkey
4. Bonial C, Corvey W, Palmer M, Petukhova V, Bunt H (2011) A hierarchical unification of lirics and verbnet semantic roles. In: 2011 fifth IEEE international conference on semantic computing (ICSC), Palo Alto. IEEE, pp 483–489
5. Bortolini U, Tagliavini C, Zampolli A (1972) Lessico di frequenza della lingua italiana contemporanea. Garzanti, Milano
6. Buitelaar P, Cimiano P, Haase P, Sintek M (2009) Towards linguistically grounded ontologies. In: Aroyo L, Traverso P, Ciravegna F, Cimiano P, Heath T, Hyvönen E, Mizoguchi R, Oren E, Sabou M, Simperl E (eds) The semantic web: research and applications. Lecture notes in computer science, vol 5554, Springer, Berlin/Heidelberg, pp 111–125
7. Burchardt A, Erk K, Frank A, Kowalski A, Pado S, Pinkal M (2006) The SALSA corpus: a German corpus resource for lexical semantics. In: Proceedings of LREC 2006, Genoa
8. Chiarcos C, Nordhoff S, Hellman S (eds) (2012) Integrating WordNet and Wiktionary with lemon. Springer, Frankfurt
9. Chiari I, Alessandro Oltramari GV (2010) Di cosa parliamo quando parliamo fondamentale? In: Atti del Convegno della Società di linguistica italiana (SLI 2010), Viterbo, pp 221–236
10. De Mauro T (1965) Introduzione alla semantica. Laterza, Bari

²⁵See for example: <http://msw2.deri.ie/>

11. De Mauro T (1980) Guida all'uso delle Parole. Editori riuniti, Roma
12. De Mauro T, Mancini F, Vedovelli M, Voghera M (1993) Lessico di frequenza dell'italiano parlato (LIP). Etaslibri, Milano
13. Fellbaum C (1998) WordNet - An Electronic Lexical Database. MIT Press, Cambridge, Massachusetts
14. Fillmore CJ (1968) The case for case. In: Bach E, Harms T (eds) Universals in linguistic theory. Rinehart and Wiston, New York
15. Gangemi A (2010) What's in a schema? A formal metamodel for ECG and FrameNet. Studies in Natural Language Processing. Cambridge University Press, Cambridge
16. Gildea D, Jurafsky D (2002) Automatic labeling of semantic roles. *Comput Linguist* 28(3):245–288
17. Giuglea AM, Moschitti A (2006) Semantic role labeling via FrameNet, VerbNet and PropBank. In: Proceedings of the COLING-ACL, association for computational linguistics, Stroudsburg, PA, USA, ACL-44, pp 929–936. DOI 10.3115/1220175.1220292
18. Kipper K, Korhonen A, Ryant N, Palmer M (2008) A large-scale classification of English verbs. *Lang Resour Eval* 42(1):21–40
19. Lenci A, Bel N, Busa F, Calzolari N, Gola E, Monachini M, Ogonowski A, Peters I, Peters W, Ruimy N, Villegas M, Zampolli A (2000) Simple: a general framework for the development of multilingual lexicons. *Int J Lexicogr* 13(4):249–263
20. Lenci A, Calzolari N, Zampolli A (2002) From text to content: computational lexicons and the semantic web. In: Eighteenth national conference on artificial intelligence; AAAI workshop, "Semantic web meets language resources", Edmonton, Alberta, Canada
21. Lenci A, Johnson M, Lapesa G (2010) Building an Italian framenet through semi-automatic corpus analysis. In: Proceedings of LREC 2010, Valletta, pp 12–19
22. Levin B (2003) English verb classes and alternation: a preliminary investigation. University of Chicago Press, Chicago
23. Màrquez L, Carreras X, Litkowski KC, Stevenson S (2008) Semantic role labeling: an introduction to the special issue. *Comput Linguist* 34(2):145–159
24. Masolo C, Gangemi A, Guarino N, Oltramari A, Schneider L (2002) WonderWeb deliverable D17: the WonderWeb library of foundational ontologies. Technical report
25. Minsky M (1997) A framework for representing knowledge. In: Winston P (ed) Mind design. MIT Press, Cambridge, pp 111–142
26. Nuzzolese AG, Gangemi A, Presutti V (2011) Gathering lexical linked data and knowledge patterns from FrameNet. In: Proceedings of international conference on knowledge capture (K-CAP2011), Banff
27. Oltramari A, Mehler A, Kühnberger KU, Lobin H, Lungen H, Storrer A, Witt A (2012) An introduction to hybrid semantics: the role of cognition in semantic resources, vol 370. Springer, Berlin/Heidelberg, pp 97–109
28. Pantel P, Bhagat R, Coppola B, Chklovski T, Hovy EH (2007) ISP: learning inferential selectional preferences. In: Proceedings of HLT-NAACL, Rochester, pp 564–571
29. Patwardhan S, Pedersen T (2006) Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In: Proceedings of the EACL 2006 workshop on making sense of sense: bringing computational linguistics and psycholinguistics together, Trento, Italy, pp 1–8
30. 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
31. Poesio M (2005) Domain modelling and NLP: formal ontologies? Lexica? Or a bit of both? *Appl Ontol* (1):27–33
32. Pradhan S, Hovy E, Marcys M, Palmer M, Ramshaw L, Weischede R (2002) Ontonotes: a unified relational semantic representation. In: Proceedings of the first IEEE international conference on semantic computing (ICSC-07), Irvine, CA
33. Prévot L, Huang CR, Calzolari N, Gangemi A, Lenci A, Oltramari A (eds) (2010) Ontology and the lexicon. Cambridge University Press, Cambridge/New York

34. Pustejovsky J (1995) *The generative lexicon*. MIT Press, Cambridge
35. Pustejovsky J, Jezek E (2008) Semantic coercion in language: beyond distributional analysis. *Ital J Linguist* 20(1):175–208
36. Reale G (1990) *A history of ancient philosophy: Plato and Aristotle*. SUNY Press, Albany
37. Resnik P (1996) Selectional constraints: an information-theoretic model and its computational realization. *Cognition* 61:127–159
38. Roventini A, Alonge A, Calzolari N, Magnini B, Bertagna F (2000) *ItalWordNet: a large semantic database for italian*. In: *Proceedings LREC 2000, Athens*, pp 783–790
39. Ruppenhofer J, Ellsworth M, Petruck M, Johnson C (2005) *FrameNet: Theory and Practice*. Available at: <http://framenet.icsi.berkeley.edu/>
40. Searle J (1995) *The construction of social reality*. Paperback. Free Press, New York
41. Settles B, Small K, Tomanek K (eds) (2010) *Proceedings of the NAACL HLT 2010 workshop on active learning for natural language processing*. Association for Computational Linguistics, Los Angeles
42. Shi L, Mihalcea R (2005) Putting pieces together: combining FrameNet, VerbNet and WordNet for robust semantic parsing. In: *Proceedings of the 6th international conference on computational linguistics and intelligent text processing, CICLing 2005*. Springer, Berlin/Heidelberg/New York, pp 100–111
43. Simons P (ed) (1987) *Parts: a study in ontology*. Clarendon Press, Oxford
44. Stevenson M, Greenwood M (2006) Learning information extraction patterns using wordnet. In: *Third international global WordNet conference (GWNC-2006)*, Jeju Island, Korea, pp 52–60
45. Suárez-Figueroa MC, Gómez-Pérez A, Motta E, Gangemi A (eds) (2012) *Ontology engineering in a networked world*. Springer, Berlin
46. Tonelli S (2010) *Semi-automatic techniques for extending the framenet lexical database to new languages*. Ph.D. thesis, Department of Language Sciences, Università Ca' Foscari Venezia, Venezia
47. 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*
48. Vetere G, Oltramari A, Chiari I, Jezek E, Vieu L, Zanzotto FM (2012) *Senso comune, an open knowledge base for Italian*. TAL – Traitement Automatique des Langues, Special Issue of the on “Free language resources”. 52(3):217–243

The People's Web Meets NLP

Collaboratively Constructed Language Resources

Gurevych, I.; Kim, J. (Eds.)

2013, XXIV, 378 p., Hardcover

ISBN: 978-3-642-35084-9