

The Lois Project: Lexical Ontologies for Legal Information Sharing

Daniela Tiscornia

Institute of Legal Information Theory and Techniques - Italian National Research Council

Abstract. Semantic metadata are expected to support search engines for legal information retrieval, providing legal knowledge to include into their search strategies. In a wide meaning, semantic metadata are 'all kind of information describing a resource'; this paper focuses on a strict notion of semantic metadata, meaning 'information about content', i.e. 'what the resource is about'. In general standard frameworks such as Dublin Core, this kind of information is expressed by the metadata 'Subject', which carries at least one preferred term from lexical resources as controlled vocabularies or encoding schemes, which are usually defined as hierarchies of terms, usually lacking a purely defined semantics. Starting from an European project recently concluded, this paper presents a methodology for building a multi-lingual semantic lexicon for law, featuring lexically and legally grounded conceptual representations, to be used either as a source of semantic metadata and as an external tool for cross lingual retrieval. The role played by semantic resources, such as lexical and formal ontologies in improving real access to legal information is analysed and the problems encountered in the methodological steps are outlined.

Keywords: Standards for legal documents, Semantic metadata, Legal ontologies

1. Access to legal content

1.1. THE ROLE OF SEMANTIC METADATA

In many countries public institutions have promoted projects aimed at improving the availability of Public Sector information on the Web and the free access of 'institutional' information. In the specific field of legal information there is a further need to join practical/technical solutions for accessing legal information¹ with a further 'social' perspective of allowing citizen to access in an 'understandable' way legal, mainly legislative data.

Today conceptual search strategies based on keywords are still missing a clear semantics of terms, and this does not allow a conceptual query expansion; therefore, there is no semantic relationship between information needs of the user and the information content of documents, apart from text pattern matching. It is necessary, therefore, to explicit

¹ Greenleaf G., *Solving the Problems of Finding Law on the Web: World Law and DIAL*, 2000(1) *The Journal of Information, Law and Technology (JILT)*.

the semantic aspects carefully so that the search is driven by a meta-description, expressed through semantic metadata, which keeps univocal references to the text, since the non-expert user has no precise idea of what he is looking for, and uses general terms of common language rather than specific legal concepts.

Semantic metadata are expected to support search engines for legal information retrieval, providing legal knowledge to include into their search strategies. In a wide meaning, semantic metadata are 'all kind of information describing a resource': this paper focuses on a strict notion of semantic metadata referring to 'information about content', i.e.: *what the resource is about*. In general standard frameworks such as Dublin Core, this kind of information is expressed by the metadata *Subject*, which carries at least one preferred term from sources as controlled vocabularies or encoding schemes, usually defined as hierarchies of terms, usually lacking a purely defined semantics. A descriptive model of contents may point out both the typologies of regulative functions and the categories of the addressees, and it would allow to overcome linguistic barriers².

Building a semantic lexicon for law is not a trivial task, as legal conceptual knowledge is closely related to language use within the legal domain, and therefore, there is, as in other terminological domains, a relatively high level of dependence between legal concepts and their linguistic realization in the various forms of legal language. The problem is even more evident in cross lingual retrieval, where corresponding terms are often absent in different languages but equivalent concepts exist in legal systems.

1.2. LEGAL LANGUAGE AND LEGAL KNOWLEDGE

When examining the legal vocabulary, we encounter two different types of semantic information associated with elements from legal text. On the one hand, there is the ontological structuring in the form of a conceptual model of the legal domain; on the other hand, there is a vocabulary of lexical items that lexicalize concepts (a lexicon), which are not necessarily restricted to the legal domain, and are associated with specific linguistic information (e.g. nouns versus verbs and syntactic preference). Therefore the conceptual model that provides the necessary structure to build a semantic lexicon needs to be integrated by a domain model so as to reflect the peculiarities of the legal domain.

² Visser P., Bench Capon T, *Ontologies in the Design of Legal Knowledge Systems; towards a Library of Legal Domain Ontologies*, Proceedings of Jurix Conference, 2000.

Law and language are connected in many ways (Sacco 2000). First of all, they have a *similar structure*: each has, at its essence, rules which are constitutive of a system and which ensure its consistency. A second aspect is the *dependency* of law on language, since regulatory knowledge must be communicated, and the written and oral transmission of social or legal rules passes through verbal expression.

In addition, legal language, like legal knowledge, has a *multi-layered structure*: according to Kalinowsky (1965), it consists of the *language of Law* and *language of Jurists*. The former is the language in which legal rules are written: not any linguistic expression in a legal text is a legal term, but every legal term is a linguistic expression. The latter is a meta-language. It is composed of a) the “judge’s language” in case law, used to speak about legal rules and about persons and behaviours bounded by legal rules; b) the “language of jurisprudence”(legal literature and legal theory), which puts legal language and judicial interpretation into concepts, to make the structure of the system consistent and systematic³.

2. Tools: lexical and formal ontologies

2.1. FROM THESAURI TO ONTOLOGIES

Restricted vocabularies are usually defined as hierarchies of terms: “The types of vocabulary for which software tools should provide support include thesauri conforming to ISO 2788, classification schemes of various types, subject heading lists, taxonomies (typically combining the hierarchical properties of classification schemes with the reciprocal relationships and other features of thesauri) and simple authority lists. [...] While a taxonomy is designed to classify things, a thesaurus is designed to help you find the right words or phrases to describe what you are ultimately looking for⁴.”

Ontologies are designed to allow computers to really interact with each other, covering all semantic metadata in its wider meaning(e.g.: publication date, authors of paper, journal publishers, publication date, normative references, etc.). On one hand, ontologies can be machine generated from good metadata; on the other, here exposed, good ontologies can enrich semantic metadata.

³ A good example is “negoziio giuridico” (juridical act): the term never appears in Italian Legislation, but is crucial in contract law to distinguish contracts from other classes of legal acts.

⁴ e-Government Unit of the UK Cabinet Office, Tomatoes are not the only fruit: a rough guide to taxonomies, thesauri, ontologies and the like, April 2005. (<http://www.govtalk.gov.uk/schemasstandards/metadata.asp>)

One relevant aspect to take in account is the *legacy* problem, as there are several semantic resources already used and distributed over the Internet, and a large amount of legal data (for instance, all the EU documents) already classified according to available thesauri and classification schemes. Re-using and harmonizing existing resources within a *reference ontology* requires to increase the precision of the semantics of the existing relations in thesauri (see sect. 4.1).

In the following it will be described the methodology for the development of a semantic framework that can be used as an extensible and flexible thesaurus in monolingual information retrieval but also as a most complex and rich *reference ontology* for concepts equivalence setting in multilingual environments and for legal knowledge conceptualization.

The building steps will be outlined according to:

- the conceptual model for building a semantic lexicon
- the ontological layers and the role of formal ontologies
- the resources harmonization
- the use of the semantic framework in knowledge representation

2.2. FORMAL ONTOLOGIES, SEMANTIC LEXICONS

A formal ontology can be considered a theory about several views (i.e. models) of reality. Formal ontologies have a multi-layered structure: *foundational ontologies* contain domain-independent concepts, relations and meta-properties, which provide ontology builders with a formal semantics, that is, formal ontological distinctions to categorize entities in a domain. A *domain ontology* is populated by concepts, relations and instances extracted in a bottom-up fashion from the domain and consistent with the top-down formal semantics imposed by the upper ontology. In complex domains such as the legal one, a *core ontology* is part of a layered architecture which intends to bridge the gaps between domain-specific concepts and the abstract categories of upper ontologies; it also expresses the basic concepts that are common across a variety of domains, providing a global and extensible model into which data originating from distinct sources or different vocabularies can be mapped and integrated (Doerr et al., 2003).

Semantic lexicons are means for content management which can provide a rich semantic repository. Compared to formal ontologies, semantic lexicons, also called *lightweight ontologies*, are generic and based on a weak

abstraction model, since the elements (classes, properties, and individuals) of the ontology depend primarily on the acceptance of existing lexical entries. In a lexical ontology, such as WordNet (Fellbaum, 1998), many of the hyper/hyponymy links might be logically inconsistent, as it was designed as a lexical resource and constraints over relations and consistency are ruled by the grammatical distinctions of language.

3. The conceptual model

The multilingual lexicon here presented is currently composed by about 35.000 concepts in five European languages. (English, German, Portuguese, Czech, and Italian, linked by English); it has been created within the European funded project LOIS (Legal Ontologies for Knowledge Sharing, EDC 22161, 2003-2006⁵) extending in a multilingual dimension a semantic lexicon for the Italian legal language, the Jur-WordNet data base.

The methodology is based on an existing de facto standard, the WordNet and EuroWord Net resources, WordNet (Fellbaum 1998) is a lexical database which has been under constant development at Princeton University. EuroWordNet (EWN) (Vossen et al., 1997) is a multilingual lexical database with wordnets for eight European languages, which are structured along the same lines as the Princeton word net (see www.globalwordnet.org).

The conceptual model makes a clear separation between a concept (meaning) and its lexicalizations (words). More precisely:

- each concept is expressed by one or more linguistic expressions (single or multi words) in the same language or in different languages;
- each word can have variant forms (singular, plural, tenses in verbs, abbreviations, etc.);
- each word can have more than one sense, i.e. can express more than one concept. Each *word sense* express one concept only.

In terms of the WordNet framework, a concept is a *synset*, the atomic unit of the semantic net. A synset is a set of one or more uninflected word forms (lemmas) with the same part-of-speech (noun, verb, adjective, and adverb) that can be interchanged in a certain context. For example, {*action, trial, proceedings, law suit*} form a noun synset because they can be used to refer to the same concept. More precisely

⁵ www.loisproject.org

each synset is a set of *wordsenses*. A synset is often further described by a gloss, explaining the meaning of the concept. English glosses drive cross-lingual linking. Synsets are linked on the basis of the following relations:

Monolingual relations

- words are linked by *lexical* relations: synonymy (included in the notion of synset and rare in the legal lexicon), near-synonym, antonym, derivation⁶
- concepts (Synsets) can be related to each other by *semantic* relations, of which the most important are *hypernymy/hyponymy* (between specific and more general concepts), meronymy (between parts or wholes), thematic *role* (between noun (agent) and verb or event- denoting nouns (actions), *instance-of*;
- synsets in the National Legal WN are (or shall be) linked by *generalization/specialization* relations to the general language modules, developed within the *EuroWordNet Project*.

Cross-lingual equivalence relations are made explicit in the so-called Inter-Lingual-Index (ILI). Each synset in the monolingual wordnets is linked by equivalence relation with an English synset, which is a record in the ILI. These relations indicate *complete equivalence*, *near equivalence*, or *equivalence as a hyponym* or *hypernym*. The network of equivalence relations (see fig.1) determines the interconnectivity of the indigenous wordnets. Language-specific synsets from different languages linked to the same ILI-record by means of a synonym relation are considered conceptually equivalent.

3.1. THE DOMAIN MODEL

To take into account the distinction between sources of law, in creating the semantic resource, a first nucleus of pilot concepts has been selected at the general level of doctrine conceptualization, offering a consolidated reference structure and allowing greater sharing. Concepts, imported from the Italian legal WordNet (JurWordNet, Tiscornia at al., 2004) have been identified within the frequency list of Italian Legislation corpus. Their selection was based on the assessment of experts. Descriptions (glosses) were extracted from legal handbooks.

⁶ Lexical relation are ruled by Part of Speech (POS) constraints (e.g. synonym and antonym holds between synset pertaining to the same POS, derivation holds between noun/verb and adjectives).

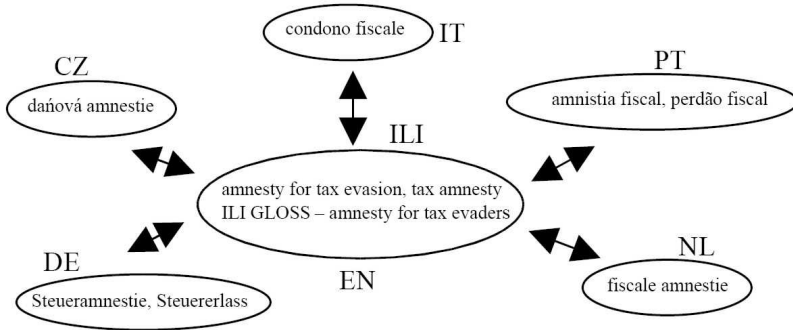


Figure 1. Example of cross lingual link through the ILI

In the second step the emphasis was put on the detection of legislative terminology, i.e. terminology that is specific to the legal domain, as opposed to the abstract conceptualization (often of common sense concepts) above. In order to identify this legislative knowledge, a parallel corpus derived from the European Directives in the consumer law domain was taken into account. Semi-automatic alignment techniques enabled the selection of a multilingual set of legal terms and the automatic generation of a unique *Identifier*. Automatic extraction of legal concepts from national legislation is limited to the consumer law domain.

A further notion of *legal concept* and proper *legal relations* had been added to the conceptual model described above:

- Legal terms were only selected if they had an explicit definition in the text: each definition is assumed to create a (*new*) *legal concept*.
- Once alignment had been established, *legal equivalence* was assumed and each set of corresponding terms in different languages were automatically linked to one unique identifier.
- an additional legal relation *implemented_as* defines the link between a European legal concept and its implementation in national legislation.
- As to legal concepts from European legislation, the *Identifier* acts as the Interlingual Index item.

The interrelation between EU and national concepts represents a special case of intra-lingual inter-modular equivalence relation, even if carried out at a mono-lingual level.

According to the two building approaches described, the main module of each national WordNet is composed of:

- an indigenous *lexical data base*, which conceptualizes general language entities pertaining to legal theory and legal dogmatic,
- a *legislative data base*, populated by concepts defined in European and national legislation and structured according to purely legal (supra) national models.

Taxonomic relations can span across the different modules (esp. lexical and national legal) which form a LOIS WordNet: legal concepts (i.e. concepts defined within legal texts) can have hypernyms - as well as near-synonyms - in the lexical database, as legal terms bear specialised meanings which might be different from the meaning of the same words within general language.

With respect to legal concepts from national legislation, the ILI can be automatically generated. If a legal concept from a European directive is implemented in indigenous legislation, and the local legal concepts are deemed (legally) equivalent to their European counterparts, then an equivalence relation between the two local concepts may also be established. In all other cases, the creation of semantic links between local synsets does not necessarily imply the creation of equivalence relations with the ILI, except when concepts from more than one indigenous WordNet coincide, in which case these will all be related to one ILI record. Within this architecture, the semantic structures peculiar to each WordNet will be preserved, and will overlap through the ILI.

External ontologies such as the DOLCE2.1-Lite-Plus + CLO (Gangemi et al., 2005) will structure the ILI concepts, classifying concepts according to explicit and consistent subsumption relations. The ontological level acts as an external ordering principle and subsumption of upper level synset have been manually performed without formalizing conceptual relations derived from legal texts. This external structure supports the management of the semantic/equivalence relations via the ILI in the integration of legislative and lexical/common sense knowledge (see figure 2 for a simplified view on the database structure).

3.1.1. *The role of external ontologies: polysemy disambiguation*

One of the most interesting functions of wordnets (which is usually missing in traditional thesauri) is the disambiguation of *polysemy*. Polysemy (one term has more than one meaning) is expressed by *word sense* numbering and by the association of one synset to each sense of a polysemic word. As an example, the Italian synset *ordine* (order) has several senses:

ordine_1, a command given either in speech or writing by a person or body having the authority to do so;

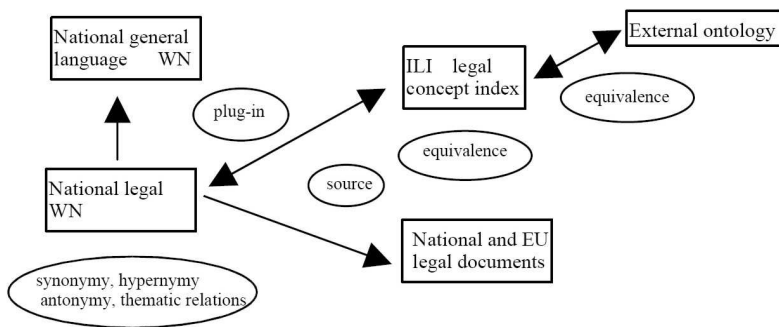


Figure 2. The overall architecture of Lois

ordine_2, a group of persons or things which form a separate/independent category, because they share a condition or some particular characteristics;

ordine_3, arrangement of separate elements according to specific criteria.

Legal polisemy

A further notion of polisemy detection, implied by the assumption of the constitutive role of legislative definitions is the consistent distinction of degrees of equivalence between contexts in which the word occurs.

The most common situation is the 'apparent polisemy' generated by the integration of the legal and lexical databases, because meanings of the legal concepts are usually more specific than the meanings of corresponding lexical items, and legal senses can display degrees of ontological overlap or even taxonomic ordering. The lexical sense can be considered to be a prototypical description of the concept properties, over which the legislator's definition impose constraint, and therefore it is classified as a hypernym of all legal senses. For instance, from EU Legislation texts, four senses of 'worker' are defined:

worker-1, any worker as defined in Article 3 (a) of Directive 89/391/EEC who habitually uses display screen equipment as a significant part of his normal work.

worker-2, any person employed by an employer, including trainees and apprentices but excluding domestic servants;

worker-3, any person carrying out an occupation on board a vessel, including trainees and apprentices, but excluding port pilots and shore personnel carrying out work on board a vessel at the quayside;

worker-4, any person who, in the Member State concerned, is protected as an employee under national employment law and in accordance with national practice.

The corresponding lexical entry is defined in the lexical part as follows: “a *person who works at a specific occupation*”.

Systematic polysemy

Often, sense distinctions do not only concern the history of language, but also interdependent notions used in organizing social reality. For instance, a typical systematic polysemy is that between an institution, a function, and a physical object: the entry *President of the Republic* can indicate a physical person, the constitutional body, or the holder of the state function. Another example, very common in Law, is the systematic polysemy involving both a normative content and a physical entity: the entry *contract* may be conceptualized as a legal transaction, as a physical document, or as an information object. The entry *appeal* can express the sense of a *petition* made to a Judge, and the *written document* of that petition. The entry *office* can express the function and the physical place where the function takes place.

While semantic lexicons can represent simple polysemy, the systematic relations occurring between the different senses cannot be expressed. Rich axiomatic theories like a Core Legal Ontology provide the needed relations to account for systematic polysemy, for example a legal transaction (e.g. the content of a contract) is *expressed* by an information object (e.g. the linguistic encoding of the content of a contract), which is *realized* by a legal document (e.g. the physical object realizing the encoding of the contract). (Gangemi, Guarino et al. 2002).

Separate classes from instances

Ontological characterization also helps the separation of classes from instances. For example, “competent authority” in the EU Directive on data protection is a class; the “*garante per la protezione dei dati personali*” in Italian legislation and the “*Agencia de Proteccion de Dato*” in Spanish legislation are instances. From a multilingual point of view, ontological classes might be even used to enhance comparison between different legal systems, by grouping similar national instantiations of a given class (e.g. the Italian “Camera dei Deputati” and the English “House of Commons” as instances of the ontological class “legal institutions”) which might not – due to the *country-dependence* of the legal domain – be perfect equivalents

3.1.2. *Formal consistency checking*

A stronger role formal ontologies could play is to check the consistency of the hierarchical structure and to distinguish language-independent concepts and relations from concepts and relations which are not; in legal term: the *core legal ontology* separates entities/concepts which belong to the general theory of law from concepts proper of national legal systems. It intends also to bridge the gap between domain-specific con-

cepts and the abstract categories of formal upper level or foundational ontologies such as DOLCE (Gangemi et al. 2002), transforming lexical relations in formal properties consistent with the top-down formal semantics imposed by the upper ontology.

To check formal consistency checking, all ontologies must be expressed in the same formal language; therefore each lexical resource has been translated in OWL format, according to the word net RDF/OWL Standard approved by the W3C Consortium (working draft June 2006). The word net schema is composed by three main classes:

- Synset: each synset is an instance belonging to one of four disjoint sub-classes (nouns, verbs, adjective and adverbs)
- wordsense: each Wordsens belong to exactly one synset
- word: each Word can be related to one or more wordsenses

Relations between synsets (including Hyponym) and relations between Wordsenses are interpreted as properties. The upper level synsets are linked by sub-class relations to the imported CLO and Dolce ontological classes, thus allowing an ontology-based consistency-checking performed by an external 'reasoner'. Some examples of the results are given below.

The main entities in DOLCE (and consequently in CLO, Core Legal Ontology) are axiomatized, disjoint classes, characterized by meta properties, such as Identity, Unity and Rigidity. As for CLO, the most relevant distinction is between *Roles* (anti-rigid) and *Types*, which are rigid. For example, every instance of a role (e.g. *student*, *plaintiff*, *guilty*) can possibly be a non student, not guilty, etc. without losing its identity. Every instance of a type (e.g. a person) must be a person. A type can play more roles at the same time. For instance, a legal subject (either a natural or artificial person) can be an owner, a tax-payer, or a murderer. In the CLO taxonomy, roles cannot subsume types, and therefore lexical concepts that are anchored to roles should not have hyponyms pertaining to types.

The core ontology also refines automatically created relations from English/ ILI records to lexicon as shown in the examples below:

- *consumer is a person, is a living thing, is a physical entity in WN; is a social role, is a non physical entity in CLO;*
- *lease is a is a contract, is a communication, is a an abstraction in WN; is a contract, is a social description, is a social concept in CLO.*

4. Expanding the lexicon

All over the Lois project development, various methodologies have been applied to populate and structure the lexicon, of which the following are the most important:

- Manual expert translation of a selected bootstrapping set of existing synsets in the Italian legal WordNet (JurWN);
- Manual creation of legal synsets on the basis of authoritative resources;
- Automatic extraction of explicitly defined concepts from legislative text (national and EU);
- Automatic extraction of significant lexical elements from legal text;
- Mapping lexical concepts onto the English WordNet and adopting its hierarchies;

To further expand the domain coverage, a bottom-up strategy will be integrated with a top-down validation. The bottom approach will be based on:

- Automatic extraction of domain concepts and dynamic integration
- Importing and harmonizing existing resources
- Axiomatizing legal definitions

In this context we will focus on the integration of external semantic resources within the structure. This requires the transformation of traditional thesauri into ontologies.

4.1. INTEGRATION WITH TRADITIONAL THESAURI

The most prominent EU thesaurus is *Eurovoc*, a multilingual thesaurus, structured as a set of independent hierarchical trees with inter-lingual relations. To import the Eurovoc *descriptors* (and *non-descriptors*) into the semantic lexicon the Thesaurus relations (*Broader Term*, *Narrower Term*, *Used For*, *Related Term*) will be translated in the WordNet semantic relations.

Tab. I shows the correspondence table.

Form a thesaurus to a WordNet lexicon, relations can be automatically translated; further manual refinements will remove misleading taxonomies, will separate instances from classes and will give a more

Table I. shows the correspondence table

| Eurovoc relations | Wordnet relations | <i>Dolce+CLO properties</i> |
|--------------------------|---|---|
| NT/BT | Hyponym/hyper/nym, instance_of, meronym | sub-class-of, instance_of, part-of, member_of |
| UF | <i>each term in a synse</i> | same_as |
| RT | role, cause | <i>ontological formal properties</i> |

connotation to the generic *Related term* relation. For instance, see in the figure 3, part of the Eurovoc microthesaurus on *ownership*.

In the taxonomy, *ownership law* (the rules regulating ownership) and *ownership right* (the normative position) are considered synonym; *acquisition of property*, *division of property*, *easement*, *expropriation*, etc. are narrower terms but not proper hyponyms (but legal *events* affecting the normative position of the owner); *land and building* are objects of possession; *law of succession* is hyponym of *ownership law*, but not of *ownership right*.

To achieve a more precise and explicit representation of terms, each descriptor must be assigned to an ontological class and proper relation

```

ownership
  NT1 acquisition of property
    RT agricultural real estate (5616)
    RT building subsidy (2846)
  NT1 division of property
  NT1 easement
  NT1 expropriation
  NT1 joint ownership
    RT multi-storey dwelling (2846)
  NT1 land and buildings
    RT agrarian law (5606)
    RT agricultural real estate (5616)
    RT building plot (2846)
    RT industrial plot (2846)
    RT price of farm land (5616)
    RT property tax (2446)
    RT real estate market (2846)
  NT2 land register
    RT building permit (2846)
    RT local tax (2446)
    RT property tax (2446)
    RT town-planning regulations (2846)
  NT1 law of succession
    RT capital transfer tax (2446)
  NT1 nationalisation
    RT mixed economy (1621)
    RT public sector (4011)
    RT public service (0436)
  NT1 personal property
  NT1 private property
  NT1 privatisation
    RT public sector (4011)
  NT1 public property
    RT publicly-owned forest (5636)
    RT State farm (5616)
    RT State-owned land (5616)

```

Figure 3. Eurovoc Thesaurus.

must be defined over the ontological classes; to give an example (Tab. II):

Table II.

| <i>Onto-class</i> | <i>Concept t(synset)</i> | <i>properties/relations</i> | <i>Onto-class</i> |
|---|--|-----------------------------|--|
| Agent (physical agent/ social individual) | natural person, organization, company | <i>participant_in</i> | Event/process |
| Role | owner | <i>played_by</i> | Agent |
| Event/process | | <i>regulated-by</i> | norm |
| Event/process | acquisition of property division of property easement expropriation | <i>affects</i> | Normative Position/ Institutional fact |
| Normative Position/ Institutional fact | right of pre-emption | <i>dependent-on</i> | norm |
| norm | law of succession | <i>defines</i> | role, social individual |

The resulting organization is shown in table III.

5. Conclusions

In this paper we have described theoretical, practical and structural aspects of the LOIS multilingual legal knowledge base. This legal knowledge repository contains legal terminology from national and European legislation within the domain of consumer law. It also holds significant lexical, general language concepts that occur in the legal documents. These concepts are interlinked within each language and between languages by means of an extended set of EWN relations.

The structure of the LOIS database allows a user to perform a concept based search for monolingual and cross-lingual legal information retrieval, which uses keywords obtained from query expansion through the structured hierarchies of the legal wordnets and the equivalence relations with the ILLI.

Furthermore, the LOIS architecture will allow users to investigate a wide range of legal research issues, such as the comparison of national legal systems through translation, equivalence and ontological structure

Table III.

| ownership, right of accession | | |
|--------------------------------------|---|-----------------------------|
| <i>sub-class</i> | personal property joint ownership, private property, public property, real property | sub-class |
| <i>affected_by</i> | transfer of property, time-sharing, access to property, privatisation nationalisation, usufruct | gift inheritance |
| <i>has_object</i> | land and buildings | agricultural real estate |
| <i>regulated_by</i> | law of property, property law, system of property law of succession right of pre-emption | |
| <i>participant</i> | owner, seller, buyer, landlord, heres | |

across the different legal wordnets, the investigation of relations between EU and national legislative documents, and an empirical inventory of the differences between common language meaning and legal meaning. The structure of the LOIS database enhances the interoperability of multilingual legal data, and allows the incremental integration of additional legal information. The role of top-level formal ontology is fundamental in this process. This ontological level not only reinforces the existing structure (polysemy detection, ILI structuring, etc.), but also assists the automatic integration of the database through ontology-building techniques.

In conclusion, the LOIS knowledge base provides a flexible, modular architecture that allows integration of multiple classification schemes, and enables their transformation in enriched semantic metadata.

References

- Doerr, M., Hunter, J., Lagoze, C., (2003) Towards a Core Ontology for Information Integration, in *Journal of Digital Information*, Volume 4 Issue 1, 2003
- Fellbaum, C. (ed.) (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L. (2002), Sweetening Ontologies with DOLCE. In: *Proceedings of EKAW 2002*.
- Kalinowsky, G. (1965). *Introduction à la logique juridique*. Pichon & Durand-Auzias, Paris.

- Gangemi, A, M.-T. Sagri, D. Tiscornia (2003). Jur-Wordnet, a Source of Metadata for Content Description in Legal Information. In: *Proceedings of the Workshop on 'Legal Ontologies & Web based legal information management', part of The International Conference of Artificial Intelligence and Law (ICAIL 2003)*, Edinburgh, June 24, 2003.
- Boris Lauser, *AGROVOC: From thesaurus to Ontology Transforming Agrovoc into a KAON ontology*, Agrovoc Workshop, Rome 2003.
- Peters W., Sagri M.T., Tiscornia D., Castagnoli S.(2006), *The LOIS Project* Lerc 2006.
- Sacco, R., Il diritto muto, in *Riv.di diritto civile*, 39 (1993), parte I, pp. 689-702.
- Sacco, R., Droit et langue. In: *Rapports italiens au XV Congrès international de droit comparé*, Milano, 1998.
- Sacco R., Lingua e diritto, in *Ars Intrepretandi*, Annuario di ermeneutica giuridica, traduzione e diritto, Milano, 2000.
- Vossen, P., Peters, W. & Díez-Orzas, P. (1997). 'The Multilingual design of the EuroWordNet Database. In: Mahesh, K. (ed.), *Ontologies and multilingual NLP, Proceedings of IJCAI-97 workshop*, Nagoya, Japan, August 23-29.