# A Development Tool For Multilingual Ontology-based Conceptual Dictionaries

**G. Ajani**[†]**, G. Boella**[∗]**, L. Lesmo**[∗]**, M. Martin**[∗]**, A. Mazzei**[∗]**, P. Rossi**[†]

[†] Dipartimento di Scienze Giuridiche, Università di Torino
V. S. Ottavio 54, 124, Torino - Italy
gianmaria.ajani@unito.it, p.rouge@inwind.it

[∗]Dipartimento di Informatica, Università di Torino
C.so Svizzera 185, 10149, Torino - Italy
{lesmo,guido,mazzei}@di.unito.it

## Abstract

This paper introduces a number theoretical and practical issues related to the "Syllabus". Syllabus is a multi-lingua ontology based tool, designed to improve the applications of the European Directives in the various European countries.

## 1. Introduction

The european union each year produces a large number of Union Directives (EUD), which are translated in each of the communitary languages. The EUD are sets of norms that have to be implemented by the national legislations. The implementation of a EUD however can not correspond to the straight transposition in a national law. An EUD is subject to further interpretation, and this process is problematic for several reasons. Here we investigate two major problems: *extra-EU polysemy* and *conceptual misalignment*.

In the case of EUD (usually adopted for harmonising the laws of the Member States), the terminological matter is complicated by their necessity to be implemented by the national legislations. In order to have a precise transposition in a national law, a Directive may be subject to further interpretation. Thus a same *legal concept* can be expressed in different ways in a Directive and in its implementing national law. The same legal concept in some language can be expressed in a different way in a EUD and in the national law implementing it. As a consequence we have an extra-EU polysemy. For example, the concept corresponding to the word *reasonably* in English, is translated into Italian as *ragionevolmente* in the EUD, and as *con ordinaria diligenza* in the transposition law.

In the EUD transposition laws a different problem arises from the different national *legal doctrines*. A legal concept expressed in an EUD can not be present in a national legal system. In this case we can talk about a conceptual misalignment. To make sense for the national lawyers' expectances, the European legal terms have not only to be translated in a sound national terminology, but they need to be correctly detected when their meanings are to refer to EU legal concepts or when their meanings are similar to concepts which are known in the Member states. Consequently, the transposition of European law in the parochial legal framework of each Member state can lead to a set of distinct national legal doctrines, that are all different from the European one. In case of consumer contracts (like those concluded by the means of distance communication as in Directive 97/7/EC, Art. 4.2), the notion to provide in a *clear and comprehensible manner* some elements of the contract by the professionals to the consumers represents a specification of the information duties which are a pivotal principle of EU law. Despite the pairs of translation in the language versions of EU Directives (i.e., *klar und verständlich* in German - *clear and comprehensible* in English - *chiaro e comprensibile* in Italian), each legal term, when transposed in the national legal orders, is influenced by the conceptual filters of the lawyers' domestic legal thinking. So, *klar und verständlich* in the German systems is considered by the German commentators referring to three different legal concepts: 1) the print or the writing of the information must be clear and legible (*gestaltung der information*), 2) the information must be intelligible by the consumer (*formulierung der information*), 3) the language of the information must be the national of consumer (*sprache der information*). In Italy, the judiciary tend to control more the formal features of the concepts 1 and 3, and less concept 2, while in England the main role has been played by the second concept, though considered as plain style of language (not legal technical jargon) thanks to the historical influences of plain English movement in that country.

To manage properly extra-EU polysemy and conceptual misalignment we distinguish in the Syllabus project the notion of legal term from the notion of legal concept and build a systematic classification based on this distinction. In Section 2 we describe the major issues of the Syllabus project, in Section 3 we relate the Syllabus approach with respect to similar projects and in Section 4 we give some conclusions.

## 2. The syllabus system

The tool that we propose is based on a clear distinction between the notions of legal term and legal concept. The basic idea is that the basic conceptual backbone consists in a taxonomy of concepts (ontology) to which the terms can refer to express their meaning. One of the main points to keep in mind is that we do not assume the existence of a single taxonomy covering all languages. In fact, it has been convincingly argued that the different national systems may organize the concepts in different ways. For instance, the term *contract* corresponds to different concepts in common law and civil law, where it has the meaning of *bargain*
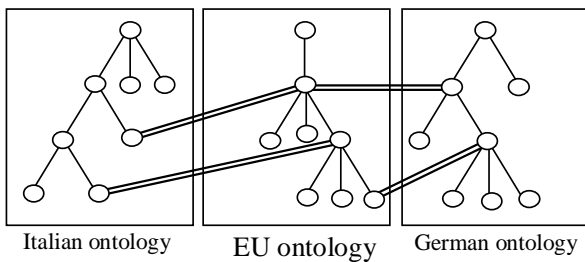
Figure 1: Relationship between ontologies. The thick arcs represent the inter-ontology "association" link.



Figure 2: Relationship between ontologies and terms.

and *agreement*, respectively (Sacco, 1999; Pozzo, 2003). In most complex instances, there are no homologous between terms-concepts such as *frutto civile* (legal fruit) and *income*, but respectively civil law and common law systems can achieve functionally same operational rules thanks to the functioning of the entire taxonomy of national legal concepts (Graziadei, 2004). Consequently, the Syllabus includes different ontologies, one for each involved language plus one for the language of EU documents. Each language-specific ontology is related via a set of *association* links to the EU concepts, as shown in Fig. 1.

### 2.1. Polysemy, conceptual misalignment and translations

Although this picture is conform to intuition, in Syllabus it had to be enhanced in two directions. First, it must be observed that the various national ontologies have a reference language. This is not the case for the EU ontology. For instance, a given term in German could refer either to a concept in the UK ontology or to a concept in the EU ontology. In the first case, the term is used for referring to a concept in the national UK legal system, whilst in the second one, it is used to refer to a concept used in the European directives. This is one of the main advantages of Syllabus. For example *klar und verständlich* could refer both to concept Ger-379 (a concept in the German Ontology) and to concept EU-882 (a concept in the European ontology). This is the Syllabus solution for facing the possibility of a correspondence only partial between the meaning a term has in the national system and the meaning of the same term in the translation of a EU directive. This feature enables Syllabus to be more precise about what "translation" means. It puts at disposal a way for asserting that two terms are the translation of each other, but just in case those terms have been used in the translation of an EU directive: within Syllabus, we can talk about direct EU-translations of terms, but only about indirect national-system translations of terms. The situation enforced in Syllabus is depicted in Fig. 2, where it is represented that: The Italian term *Term-Ita-A* and the German term *Term-Ger-A* have been used as corresponding terms in the translation of an EU directive, as shown by the fact that both of them refer to the same EU-concept EU-1. In the Italian legal system, *Term-Ita-A* has the meaning Ita-2. In the German legal system, *Term-Ger-A* has the meaning Ger-3. The EU translations of the directive is correct insofar no terms exist in Italian and German that characterize precisely the concept EU-1 in the
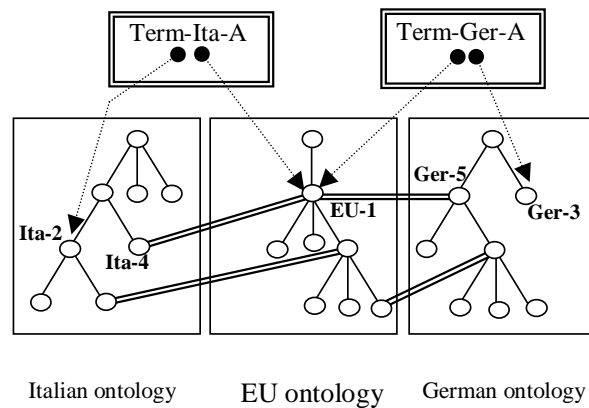
two languages (i.e., the "associated" concepts Ita-4 and Ger-5 have no corresponding legal terms). A practical example of such a situation is reported in Fig. 3, where we can see that the ontologies include different types of arcs. Beyond the usual *is-a* (linking a category to its supercategory), there are also a *purpose* arc, which is self-explanatory, and *concerns*, which refers to a general relatedness. The dotted arcs represent the reference from terms to concepts. Some terms have links both to a National ontology and to the EU Ontology (In particular, *withdrawal* vs. *diritto di recesso* and *difesa del consumatore* vs. *consumer protection*).

The last item above is especially relevant: note that this configuration of arcs specifies that: 1) *withdrawal* and *diritto di recesso* have been used as equivalent terms (concept EU-2) in some European Directives (e.g., Directive 90/314/EEC). 2) In that context, the term involved an act having as purpose the some kind of protection of the consumer. 3) The terms used for referring to the latter are *consumer protection* in English and *difesa del consumatore* in Italian. 4) In the British legal system, however, not all *withdrawals* have this goal, but only a subtype of them, to which the code refers to as *cancellation* (concept Eng-3). 5) In the Italian legal system, the term *diritto di recesso* is ambiguous, since it can be used with reference either to something concerning the *risoluzione* (concept Ita-3), or to something concerning the *recesso* proper (concept Ita-4). All of this seems to correspond neatly to the conception of terminology that is currently accepted by the scholars in comparative law. For instance, it can safely be stated that the term *diritto di recesso* as used in consumer law directives (i.e., the right of a consumer to withdraw from a contract) does not correspond to the same legal concept in the Italian legislation (i.e., Art. 1373 c.c.). Moreover, the right of withdrawal appearing in EU directives also differs from the term with the same label of the British system. Particularly, the EU right of withdrawal (for the lack of providing information to consumers or for an incorrect information) appears to be both an incomplete definition of a right and an indeterminate remedy, where in some remedial legal order (e.g. English) it is difficult to accept the notion of such a right, but easier to apply it in the courts as a flexible type of cancellation or, alternatively, termination of contract; while in substantive rights based orders (Italy) it is easy to en-
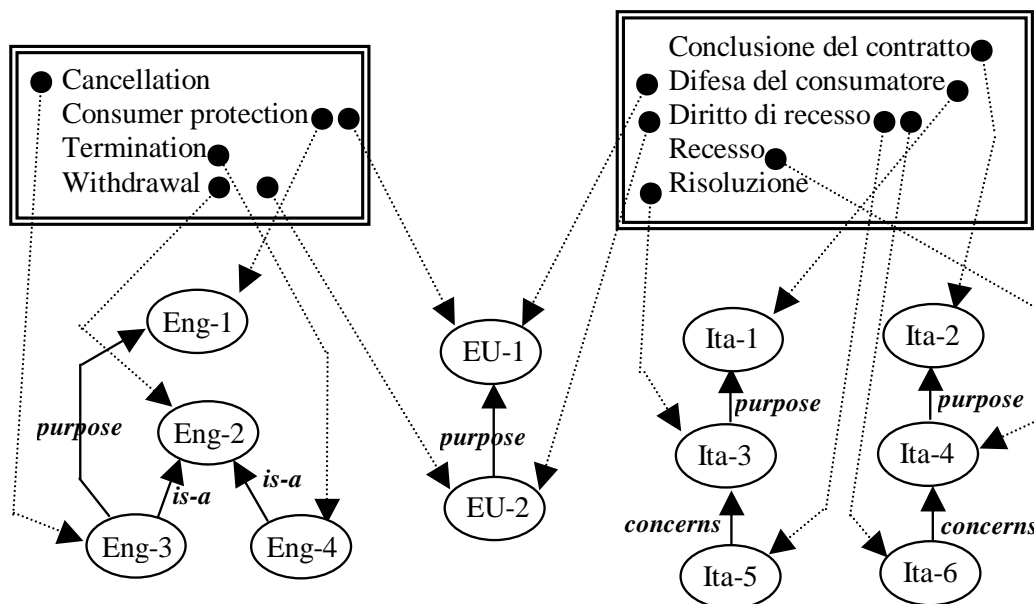
Figure 3: An example of interconnections among terms.

shrine a new right, but more difficult to enforce it beyond the classical remedy of contractual law. From the standpoint of the Syllabus, the legal concepts have been sorted in three different classification schemes: notions, rights and remedies. The analysis of such taxonomy is not presented here.

## 2.2. Top-down and bottom-up concept annotation

This complex scenario shows how the traditional top-down approach to the development of legal ontologies (Visser and Bench-Capon, 1998) is not flexible enough. Usually, ontologies are built starting from very general concepts which are then specialised in more detailed concepts. Moreover most ontologies are oriented to a single national tradition. In this process the knowledge engineers risk not to take into account the interpretation process of the legal specialists on the real multilingual data. These ontologies aim at modelling the legal code but not the legal *doctrine*, that is the work of interpretation and re-elaboration of the legal code which is fundamental for transposing EUD into national laws. In the development of the ontologies described in the previous section, we propose the two-step procedure pursued in the *UT project* (Ajani and Ebers, 2005; Rossi and Vogel, 2004). UT project (Uniform Terminology For European Private Law) is a Research Training Network (RTN) funded by European Commission (Contract n. HPRN-CT-2002-00229). The research network involves researchers from 7 universities spread across England, France, Germany, Italy, Netherlands, Poland, Spain. The results achieved by the Network can be divided between those relating to a better understanding of the historical divergences hampering uniform terminology, and those relating to the promotion of a common terminology in EU private law.

As a first step, terms are collected in a database together with the legal sources where they appear, in order to identify the concepts. Then, for each different ontology (i.e.,

each specific language ontology and the general EU ontology), the set of concepts is organized in an ontology which can be different for different legal traditions. This reconstruction work is done by legal experts rather than knowledge engineers. In this phase the result is a lightweight ontology rather than an axiomatic one. Only relations among terms are identified without introducing restrictions and axioms. The function of these ontologies is to compare the taxonomic structure in the different legislations, to provide a form of intelligent indexing and to draw new legal conclusions. In a second phase, a knowledge engineer can reorganize the ontology and integrate it with a top-level well-founded ontology like DOLCE (Gangemi et al., 2002). This is a very complex step, since the legal concepts are linked to concepts related to our everyday life in a rather intricate way (Boella et al., 2006). But it is important to note that this second step affects only in part the proper function of Syllabus, that can be used as an access tool (see below) before the integration step is completed.

## 2.3. Implementation features

Even if many tools for the construction of ontologies are available (e.g., Protegè), we had to design a new development system based on the constraints of being *distributed* and *user-friendly*. As described in Section 2.2., the central step of the legal concepts annotation is performed by legal experts from different countries rather than knowledge engineer. As a consequence, the Syllabus has been designed by using a very simple client-server web application[1]. The ontology framework has been inspired by the Gene Ontology project (http://www.geneontology.org/), from which it inherits the logical and graphical representation, and the ability to import and export the ontology in representation languages like OWL. The tool has two levels of use.

---

[1]The web server has been implemented by using a WAPP (Windows, Apache, PostgreSQL, PHP) platform.

Figure 4: A screenshot of the Syllabus interface for navigation and update of the ontologies.

In the first level the web interface makes available to the legal expert a friendly way to introduce data about terms and concepts. At this level, the tool aims at providing the expert with a support in her/his activity of term comparison, Fig. 4. In other words, instead of using a standard database interface, the expert can specify the correspondences among terms found via the manual inspection of EU directives or ECJ decisions (or in national legislation and case-law) in a controlled way and save in an ontology structure the result of her/his analysis.

The second level is devoted to a user who wants to retrieve the documents related to a given legal term. At this level, Syllabus acts as the desired extension of a standard legal database (cf. "EUR-Lex", see below), by enabling the user to find the relevant documents taking into account the complex net of semantic correspondences that characterize the relationships between legal terms at the international level (Fig. 5).

The importance of such two levels of use has ever not been stressed sufficiently. As pointed out in jurisprudence and European law literature (Patterson, 1996; Gerven, 2000), the classification schemes of legal objects (such as the distinction between rights *in rem* and rights *in personam* with the reference to the time-share property) does not exist externally to the legal domain and exists only because the legal science deems it does. The principle of consensus (the intentional use of the majority of scholars operating within the relevant discipline) and not only the interoperability may foster the ontological knowledge at the level of legal domain. Consequently, if many classification schemes are adopted in several legal orders the ontology should take

all into account before refining them.

## 3.  Related work

There is a number of works that consider the theoretical issues related to the construction of legal ontologies (Mc-Carty, 1989; Stamper, 1991; Breuker et al., 1997). In particular the framework presented in (Kraligen, 1997) is a frame-based system that classify the legal facts. A basic component of this system is the *legal concept description*, i.e. Kralingen proposes a distinction between a legal term and a legal concept similar to the distinction that we have adopted in Syllabus.

From a practical point of view, there are two projects that are related in someway to the Syllabus. The "EURLex" system (http://europa.eu.int/eur-lex/) is a web portal that interfaces a number of databases in order to access a wide collection of legal documents produced by the EU. However, in order to obtain a full coverage, EURLex limits the complete accessibility to legal documents, particularly for the needs of lawyers. Each query, even when using boolean search, reports too large instances without comprehensible classifications for the expectances of national jurists and practitioners, and thus hinders the applicability of EUR-Lex for most legal uses in the Member States' legal.

"Eurovoc" (http://europa.eu.int/celex/eurovoc/) is a web application that accesses a number a multilingual thesauri. The main point of this project is the splitting of the legal terms into two sets: the *descriptor* and *non-descriptor*. A non-descriptor legal term can be always be mapped into a descriptor legal term that has the same meaning. Moreover, the basic hypothesis is that each descriptor can be translated

Figure 5: A screenshot of the Syllabus interface for searching a legal term.

straightforwardly into the official languages of the EU. In contrast to Syllabus, the main purpose of Eurovoc is the information extraction. Indeed, the sparsness problems related to the *bags of word* techniques can be reduced by replacing the non-descriptor with the corresponding descriptor. However Eurovoc does not distinct between a legal terms and a legal concepts, and cannot resolve easily the problems related to the polysemy.

"LOIS" Project (http://www.loisproject.org) aims at extending EuroWordnet with legal information. Whilst the final goal of LOIS is to support applications concerning information extraction, the Syllabus we propose herein is concerned with the access of human experts to the EU documents.

## 4. Conclusion and development

In this paper we have described the Syllabus. The first ongoing phase of this project involves the collections of the legal terms and legal concepts by a group of legal experts belonging to a number different countries (in particular Italy, England, Germany). Moreover these groups are organizing the ontologies by taking into account the national doctrines. In the second phase of the project, a group knowledge engineer will reorganize the ontology and integrate it with a top-level well-founded ontology like DOLCE (Gangemi et al., 2002).

## 5. Acknowledgments

## 6. References

G. Ajani and M. Ebers, editors. 2005. *Uniform Terminology for European Contract Law*. Nomos, Baden Baden.

G. Boella, L. Lesmo, and R. Damiano. 2006. On the ontological status of plans and norms. *Artificial Intelligence and Law*, ??:??–?? In press.

J. Breuker, A. Valente, and R. Winkels. 1997. Legal ontologies: A functional view. In *Procs. of 1st LegOut Workshop on Legal Ontologies*, pages 23–36.

A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. 2002. Sweetening ontologies with dolce. In *Proceedings of EKAW*, pages 21–29.

Van Gerven. 2000. Of rights, remedies and procedures. *Common Market Law Review*, pages 501–536.

M. Graziadei. 2004. Tuttifrutti. In P. Birks and A. Pretto, editors, *Themes in Comparative Law*, pages –. Oxford University Press.

V. Kraligen. 1997. A conceptual frame–based ontology for the law. In *Procs. of 1st LegOut Workshop on Legal Ontologies*, pages 15–22.

L.T. McCarty. 1989. A language for legal discourse: Basic features. In *Proc. of Second International Conference on Artificial Intelligence and Law*.

D. Patterson. 1996. *Law and Truth*. Oxford University Press.

B. Pozzo. 2003. Harmonisation of european contract law and the need of creating a common terminology. *European Review of Private Law*, 6:754–767.

P. Rossi and C. Vogel. 2004. Terms and concepts; towards a syllabus for european private law. *European Review of Private Law (ERPL)*, ?(?):293–300.

R. Sacco. 1999. Contract. *European Review of Private Law*, 2:237–240.

R.K. Stamper. 1991. The role of semantics in legal expert systems and legal reasoning. *Ratio Juris*, 4(2):219–244.

P.R.S. Visser and T.J.M. Bench-Capon. 1998. A comparison of four ontologies for the design of legal knowledge systems. *Artificial Intelligence and Law*, 6:27–57.