

Legal Taxonomy Syllabus version 2.0

Gianmaria Ajani¹, Guido Boella², Leonardo Lesmo², Marco Martin²,
Alessandro Mazzei², Daniele P. Radicioni², and Piercarlo Rossi³

¹ Dipartimento di Scienze Giuridiche - Università di Torino

² Dipartimento di Informatica - Università di Torino

³ Dipartimento di Studi per l'Impresa e il Territorio - Università del Piemonte
Orientale

gianmaria.ajani@unito.it, {guido,lesmo,mazzei,radicion}@di.unito.it,
notmart@gmail.com, piercarlo.rossi@eco.unipmn.it

Abstract. The need for managing the conceptual representation of European law led to the development of the Legal Taxonomy Syllabus (LTS) and the related methodology. In this paper we consider further legal issues that emerged during the test and use phases, and outline the new features that we added to the new version, the LTS 2.0.

1 Introduction

European Union Directives (EUDs) are sets of norms that have to be implemented by the national legislations and translated into the language of each Member State. The general problem of multilingualism in European legislation has recently been tackled by linguistic and ontological tools [8,5,17,18]. The management of EUD is particularly complex, since the *implementation* of a EUD does not correspond to a straight transposition into a national law.

In previous work we carried out the Legal Taxonomy Syllabus⁴ (LTS), a tool to build multilingual conceptual dictionaries aimed at representing and analysing terminologies and concepts from EUDs [1,2]. LTS is based on the distinction between *terms* and *concepts*. The latter ones are arranged into ontologies that are organised in levels. Only two levels were defined: the European level –containing only one ontology deriving from EUDs annotations–, and the national level –hosting the distinct ontologies deriving from the legislations of EU member states.

While annotating the EUDs, testing and using the system, some more requirements emerged from users expert in law, demanding for a more sophisticated approach along with further developmental efforts: first, it is frequent the case of concepts which are the result of a doctrinal interpretation process rather than of the definition in directives. If, on the one hand, the definitions in directives and their relation with the actual text are required by legal scholars to have a precise model of European law, the layman is more interested in the

⁴ <http://www.eulawtaxonomy.org>

concepts which results from the doctrinal interpretation. Furthermore, laws are typical objects evolving through time. An open issue to cope with in building legal frameworks both at the European and at the national level is the *normative change* [12,4]. Concepts in the legal ontologies should not only represent the consolidated legal text, but should also keep trace of the evolution of meaning.

In this paper we consider not only the terms defined in the directives, but also the interpretation process of legal scholars in the LTS and how to better integrate concepts and the text of EUDs in the LTS. We answer the first question by introducing *abstract* concepts (*abstract* in that they are not related to a single directive), which should be conveniently recognized as a *grouping* of concepts. The users will be thereby allowed to navigate the ontology at different levels of details depending on their goals. Moreover, exploiting natural language processing techniques we greatly simplify the management of legal text associated to concepts. Also, we investigate how to extend the ontology with a temporal dimension to the ends of representing *normative change*, and to allow users to search also for past meanings of terms and the modified norms introducing them. To these ends, we introduce *time* into the ontology, and allow new concepts to replace the old ones while keeping the latter ones in the system as well.

2 Multilingual and Multilevel Ontologies for European Directives

Comparative Law has identified two key points in dealing with EUD, which make more difficult dealing with the polysemy of legal terms. We call them the *terminological* and *conceptual misalignments*. The first problem is determined by the lexical ambiguity of the legal terms (in particular homonymy) in the translation of EUDs. The second problem is determined by the lexical and conceptual ambiguity of the legal terms (in particular polysemy) in the implementation of EUDs. These issues determined the development of the first release of the LTS, and have been illustrated in [2].

We now illustrate further issues in handling EUDs that required to devise further features to enrich the original LTS.

2.1 Concepts Abstraction

The LTS system relies on the concept of *unitary-meaning* or *umeaning*: such atomic concepts can be derived from excerpts of the text of legal norms, such as European directives or national laws, and are arranged into two separate categories of *umeanings*, as described in [2]. EUDs provide rigorous definitions of some terms, such as the definition of the Italian term *consumatore* (*consumer*), in the Italian version of the *EUD 93/13/EEC*, Art. 2 is:

[...](b) “consumatore”: qualsiasi persona fisica che, nei contratti oggetto della presente direttiva, agisce per fini che non rientrano nel quadro della sua attività professionale; [...]

[...] (b) “consumer”: means any natural person who, in contracts covered by this Directive, is acting for purposes which are outside his professional activity; [...] (*our literal translation*)

However, two facts must be pointed out. Different EUDs might affect different aspects of the legislation: thus the definition of a term in a EUD only applies to a specific context. Furthermore, EUDs could be written at different points in time, and they can introduce diverging definitions. Let us consider the definition of *consumatore*, as it appears in the Italian version of the *EUD 2002/65/EC*, *Art. 1*:

[...] (d) “consumatore”: qualunque persona fisica che, nei contratti a distanza, agisca per fini che non rientrano nel quadro della propria attività commerciale o professionale; [...]

[...] (d) “consumer”: means any natural person who, in distance contracts covered by this Directive, is acting for purposes which are outside his business or professional activity; [...] (*our literal translation*)

We remark that in contrast with English, in Italian the second definition of *consumatore* is broader than the first one, since the term *professionale* (*professional*) does not include *commerciale* (*business*). This divergence of term definitions can often occur, since EUDs have usually a sectorial specific target. In this way, EUDs covering different sectors can provide different definitions, and as many views on the same concept. Lawyers and legislators started to put together highly sectorial concepts into more abstract concepts with broader meaning, in order to describe (complex) entities, such as the *consumatore* in all of its aspects.

In recent years, in the Italian legislation EUDs are not being implemented as single laws, but rather as groups of EUDs. The juridical concepts are defined as the union of all the sectorial concepts provided by the individual EUDs, as a result of the doctrinal interpretation process of directives. These problems are common to all European languages. Consider, for instance the definition of *consumer*, in the English version of the *EUD 1999/44/EC*, *Art. 1.2* is:

[...] (a) consumer: shall mean any natural person who, in the contracts covered by this Directive, is acting for purposes which are not related to his trade, business or profession; [...]

that has a different meaning with respect to the definition of *consumer* given in the Council Directive 90/314/EEC, *Art. 2.4*:

[...] “consumer” means the person who takes or agrees to take the package (‘the principal contractor’), or any person on whose behalf the principal contractor agrees to purchase the package (‘the other beneficiaries’) or any person to whom the principal contractor or any of the other beneficiaries transfers the package (‘the transferee’) [...]

The LTS should be able to represent both the more specific dimension related to the definitions in EUDs and the more abstract one which results from the

doctrinal interpretation of European law. The LTS allows inserting the text paragraphs where *umeanings* are defined. However, to gain better understanding of legal concepts, it is often required to consider a broader fragment. For example, in the case of *consumer* the definition is not enough, and it is necessary to collect multiple paragraphs where consumer protection norms are presented and discussed.

2.2 Normative Change

Another big open issue to cope with in building tools for describing legal frameworks both at the European and at the national level is the *normative change* [12]. One major problem, well-known in the literature, is the update of *non-monotonic* ontologies and knowledge bases [4]. In other words, not necessarily ontologies and knowledge bases have a structure constant through time (e.g., see [16]): concepts and relations present in the ontology can become obsolete as new concepts and relations are added. This is indeed the case of legal frameworks, that are continuously modified as new laws can modify paragraphs of old ones.

We can have two types of normative change: *explicit* change and *implicit* change. In the first case the new norm explicitly states the abrogation of a specific paragraph of an old law (for details on this line of investigation, please refer to [6]). Alternatively, the newer law can state a concept in contradiction to previous laws, but without mentioning them explicitly. In this case the concept stated by the new law becomes the current one; also, the parts of the old laws affected by changes (no longer updated) become obsolete.

3 From LTS 1.0 to LTS 2.0

In this Section we first summarize the functionalities of the existing LTS [2], and then we explain how it has been extended to cope with the new requirements described in the previous Section.

3.1 LTS 1.0

The main assumptions of our methodology come from studies in comparative law [13] and ontologies engineering [10]. Terms –*lexical entries* for legal information–, and concepts must be distinguished; for this purpose we use lightweight ontologies, i.e. simple taxonomic structures of primitive or composite terms together with associated definitions. They are hardly axiomatized as the intended meaning of the terms used by the community is more or less known in advance by all members, and the ontology can be limited to those structural relationships among terms that are considered as relevant.

We distinguish the ontology implicitly defined by EUD, the *EU level*, from the various national ontologies. Each one of these “particular” ontologies belongs to the *national level*: i.e., each national legislation refers to a distinct national legal ontology. We do not assume that the transposition of an EUD automatically

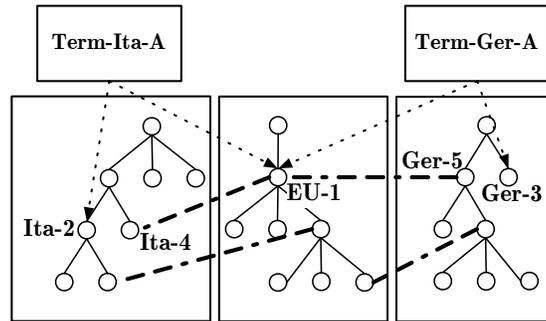


Fig. 1. Relationship between ontologies and terms. The thick arcs represent the inter-ontology “association” link.

introduces in a national ontology the same concepts that are present at the EU level.

Corresponding concepts at the EU level and at the national level can be denoted by different terms in the same national language.

A standard way to properly manage large multilingual lexical databases is to make a clear distinction among terms and their interlingual acceptions (or *axies*) [15].

In the LTS project to properly manage terminological and conceptual misalignment, we distinguish the notion of *legal term* from the notion of *legal concept* and we build a systematic classification based on this distinction. The basic idea in our system is that the conceptual backbone consists in a taxonomy of concepts (ontology) to which the terms can refer in order 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, 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* and *agreement*, respectively [14]. 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 similar operational rules thanks to the functioning of the entire taxonomy of national legal concepts [9]. Consequently, the LTS includes different ontologies, one for each involved national 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.

Although this picture is conform to intuition, in the basic LTS it has been implemented by taking two issues into account. 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 English 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,

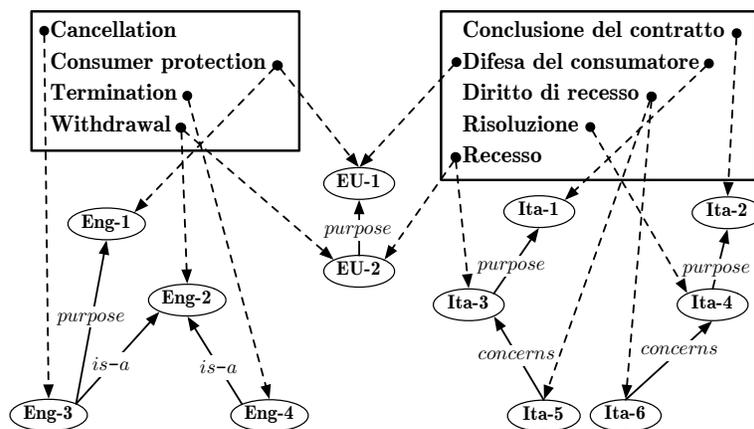


Fig. 2. An example of interconnections among terms.

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 LTS. 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 LTS solution for facing the possibility of a partial correspondence between the meaning of a term in the national system and the meaning of the same term in the translation of a EU directive. This feature enables the LTS to be more precise about what “translation” means. It makes available 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 LTS, we can talk about direct EU-to-national translations of terms, and about *implicit* national-to-national translations of terms. In other words, we distinguish between *explicit* and *implicit* associations among concepts belonging to different levels. The former ones are direct links that are explicitly used by legal experts to mark a relation between concepts. The latter ones are indirect links: if we start from a concept at a given national level, by following a direct link we reach another concept at European level. Then, we will be able to see how that concept is mapped onto further concepts at the various national levels.

The situation enforced in LTS is depicted in Fig. 1, 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 situa-

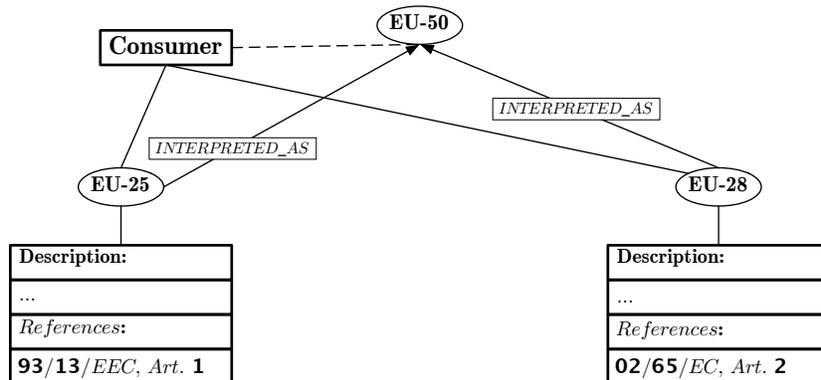


Fig. 3. Umeanings Eu-25 and Eu-28 are interpreted by the more abstract umeaning Eu-50, the link between Eu-50 and the term “consumer” is implicit.

tion is reported in Fig. 2, 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 the arcs *purpose*, which relate a concept to the legal principle motivating it, and *concerns*, which refer 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. *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 *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 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-4), or to something concerning the *recesso* proper (concept Ita-3).

3.2 Enhancing LTS with interpretation and abstraction

As described in Section 2.1, different pieces of legislations can bear different definitions of terms. Having different detailed definitions is important during the interpretation of very sectorial legal cases, but for the general case it is important to have a view that abstracts from the peculiarities of specific domains.

In order to solve this problem we introduced a new kind of ontologic relation called *INTERPRETED_AS*: it is a non transitive relation where the more general umeaning, that we will call *group leader* represents the abstracted concept that

groups the meaning of a number of more specific umeanings, that are the sectorial umeanings defined in the individual EUDs or national laws (see Fig. 3).

We have also introduced a number of constraints and integrity checks to ensure that the semantics of the grouping concept is respected and to improve the usability of the system: *i*) each umeaning can belong to a single group; *ii*) a group leader cannot exist without group members; *iii*) when the user searches into the umeaning database, more specific umeanings are excluded from the results unless the user explicitly asks to show them, i.e. only the group leaders are shown in the results. The need to contextualize concepts to the EUDs defining them leads to the need of more complex instruments to deal with the language of the norms. An umeaning is defined by the legal texts themselves; this makes clear that the creation of umeaning is a quite long task, because it requires from the user searching and reading a very large number of documents.

In order to ease this process, we developed a database that contains the full versions of the desired EUDs and national laws. In this way, the user can carry out his task according to the following workflow. 1) The user creates a new *umeaning* linked with the term he wants to define; 2) He selects relevant citation from legal text; consequently, the browser is redirected to a search page and the main term attached to the umeaning is used as the default query; 3) After choosing one of the search results, the full text of the legal document is displayed, with the search terms highlighted; 4) Finally, the user selects the text that will go in the citation with the mouse and confirms the insertion in the references database. Lastly, when the user searches for a term in the documents database, the search is not performed upon the exact words, rather with their roots, so for instance when performing the search on the term “contracts” also documents containing only “contract” will be found, this seems to enhance the information retrieval performance as shown in [11].

3.3 LTS with normative change

When a new normative is approved and enacted it can define a number of new umeanings; moreover it can happen that the same law can change a number of old umeanings defined by old laws. In particular, these old umeaning can become obsolete and no longer valid. We are aware of the difficulties concerning the modelling of the time in artificial intelligence and in formal ontology too⁵. Anyway, in LTS we adopted a naive solution in order to manage the simpler situation concerning t In the LTS it was necessary to delete all old umeanings, causing the loss of all historic informations from the database, informations that are quite valuable to better understand the evolving of the normative. This problem was resolved by using the same solution adopted for the interpretation and abstraction of the norms (Section 3.2), i.e. empowering LTS with a new ontological relation called *REPLACED_BY*.

When the paragraph of an EUD defining an umeaning has been modified by a new EUD, the new one defines a new umeaning that will replace the old umean-

⁵ E.g. see [3] for a general survey and [12] for normative systems

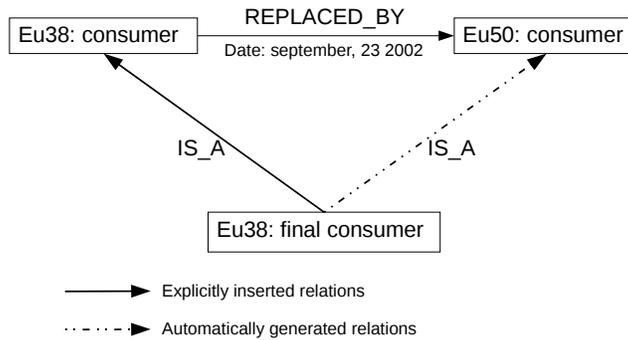


Fig. 4. An example of use of the *REPLACED_BY* relation.

ing in the ontology. There will be a relation of type *REPLACED_BY* between the two umeanings, where the child umeaning is replaced by the more general umeaning. Also in this case the new ontological relation has some peculiar characteristics that distinguishes it from the usual ontological relations (Figure 4): *i*) a *REPLACED_BY* relation brings with it a new data field not present in the other relations: the substitution date; *ii*) when the user performs a search in the umeanings database the replaced ones will not be shown, unless the user asks for a certain past date, thus obtaining a snapshot of the legal ontology that was valid in that particular moment; *iii*) when a new umeaning replaces an old one all the ontological relations where the old umeaning appeared are automatically copied in the new umeaning. If some of them are no longer valid with the new umeaning, manual intervention from the user is required.

4 Conclusions

In this paper we discuss some features that have recently been introduced in the LTS, a tool for building multilingual conceptual dictionaries for the EU law. The tool is based on lightweight ontologies to emphasize the distinction between concepts and terms. Different ontologies are built at the EU level and for each national language, to deal with polysemy and terminological and conceptual misalignment.

The present work illustrates how to distinguish between concepts as they are defined in the text of the directives and the concepts representing the doctrinal interpretation of the terms. Moreover, we point out how to deal with normative change by introducing a temporal dimension in ontologies.

Future work will involve exploring how to extend the LTS ontology, with special focus on the issue of populating it at the various levels by semi-automatic approaches [7].

References

1. G. Ajani, G. Boella, L. Lesmo, M. Martin, A. Mazzei, and P. Rossi. A development tool for multilingual ontology-based conceptual dictionaries. In *Proc. of 5th International Conference on Language Resources and Evaluation, LREC06*, pages 1–6, Genoa, 2006.
2. G. Ajani, G. Boella, L. Lesmo, A. Mazzei, and P. Rossi. Terminological and ontological analysis of european directives: multilinguism in law. In *11th International Conference on Artificial Intelligence and Law (ICAIL)*, pages 43–48, 2007.
3. J. F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154, 1984.
4. M. Cadoli and F. M. Donini. A Survey on Knowledge Compilation. *AI Communications*, 10(3–4):137–150, 1997.
5. P. Casanovas, N. Casellas, C. Tempich, D. Vrandečić, and R. Benjamins. OPJK modeling methodology. In *Proceedings of the ICAIL Workshop: LOAIT 2005*, 2005.
6. M. Cherubini, G. Giardiello, S. Marchi, S. Montemagni, P. Spinosa, and G. Venturi. NLP-based metadata annotation of textual amendments. In *Proc. of WORKSHOP ON LEGISLATIVE XML 2008, JURIX 2008*, 2008.
7. P. Cimiano. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer, 2006.
8. S. Després and S. Szulman. Merging of legal micro-ontologies from european directives. *Journal of Artificial Intelligence and Law*, February 2007.
9. M. Graziadei. Tuttifrutti. In P. Birks and A. Pretto, editors, *Themes in Comparative Law*, pages –. Oxford University Press, 2004.
10. M. Klein. Combining and relating ontologies: an analysis of problems and solutions. In *Workshop on Ontologies and Information Sharing, IJCAI'01*, Seattle, USA, 2001.
11. R. J. Krovetz. *Word sense disambiguation for large text databases*. PhD thesis, University of Massachusetts, 1995.
12. M. Palmirani and R. Brighi. Time model for managing the dynamic of normative system. *Electronic Government*, pages 207–218, 2006.
13. P. Rossi and C. Vogel. Terms and concepts; towards a syllabus for european private law. *European Review of Private Law (ERPL)*, 12(2):293–300, 2004.
14. R. Sacco. Contract. *European Review of Private Law*, 2:237–240, 1999.
15. G. Sérasset. Interlingual lexical organization for multilingual lexical databases in NADIA. In *Proc. COLING94*, pages 278–282, 1994.
16. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics* <http://genetics.nature.com>, 25:25–29, 2000.
17. Daniela Tiscornia. The Lois Project: Lexical Ontologies for Legal Information. In *Proceedings of the V Legislative XML Workshop*. European Press Academic Publishing, 2007.
18. P. Vossen, W. Peters, and J. Gonzalo. Towards a universal index of meaning. In *Proc. ACL-99 Siglex Workshop*, 1999.