Construction of a Legal Ontology from a European Community Legislative Text

Sylvie Despres¹, Sylvie Szulman²

¹ CRIP5 Equipe IAA, Univ. Paris 5 ² LIPN, UMR7030 Univ. Paris 13

¹ sd@math-info.univ-paris5.fr
² ss@lipn.univ-paris13.fr

Abstract. This paper describes the construction of an application ontology around the concept of EMPLOYEE in a European community legislation text. Firstly, the TERMINAE method is presented. Then it is shown how the text contributes to the building of this ontology. Afterward the use of a top level ontology as DOLCE and of a legal ontology as LRI-Core are detailed. The essential points of this study are: the application of the TERMINAE method to the legal domain and its alignment to core ontologies.

1 Introduction

This paper presents the construction of a legal ontology from a European community legislation text. This construction is achieved both by building the ontology from texts by using the semi-automatic TERMINAE method [1] and aligning it with a top-level ontology. TERMINAE is based on knowledge elicitation from text and allows creating a domain model by analyzing a corpus with NLP tools. The method combines knowledge acquisition tools based on linguistics with modeling techniques so as to keep links between models and texts. During the building process [2], it is assumed that: (1) the ontology builder should have a comprehensive knowledge of the domain, so that she/he will be able to decide which terms (nouns, phrases, verbs or adjectives) are domain terms and which concepts and relations are labeled with these domain terms; (2) the ontology builder knows well how the ontology will be used. The alignment process takes place during the construction. [3] defined ontology alignment as follows: ontology alignment consists in establishing links between ontologies and allowing the aligned ontologies to reuse information one from the other. In alignment, the original ontologies persist, with links established between them. Alignment usually is performed when the ontologies cover complementary domains. Our ontology is structured around two central ontologies DOLCE [4] and LRI-Core [5]. The resulting ontology does not become part of the DOLCE ontology but uses its top-level distinction. The process of ontology alignment was carried out during the ontology construction and was performed mostly by hand, with the TERMINAE tool. TERMINAE provides easy import of concepts among DOLCE but doesn’t check whether consistency is maintained after the performed operations.

The article is structured as follows: first different legal ontologies and their associated methods of construction are recalled. Then the semi-automatic TERMINAE method of ontology construction from texts, TERMINAE, is described. In the last section, the first results of the ontology construction are presented.

2 Legal ontologies

For more than a decade, legal ontologies have been developed. These ontologies were constructed for various projects concerned with the development of legal knowledge system
and legal information management. Among the best known legal ontologies the following ones can be mentioned: FOLaw (Functionnal Ontology of Law) [6] and [7] followed by LRI-Core [5]; Frame-Based Ontology [8] and [9] and more recently CLO (Core Legal Ontology) and Jurwordnet [10]; and the documentary ontology of French law dedicated to information retrieval on the Web [11]. The core ontology FOLaw, had different purposes: one of them was to distinguish the various types of knowledge in legal reasoning, and in particular the ones that are typical for reasoning. The major types of knowledge distinguished are normative knowledge, world knowledge, responsibility knowledge, creative knowledge and meta-legal knowledge. Currently [5] consider FOLaw as the inference structure of a very large solving problem (psm) for law and therefore can be added to the Commonkads [12] library of PSMs. Because FOLaw was not sufficiently detailed and was more an epistemic framework, [5] developed a new approach and constructed the LRI-Core core ontology. LRI-Core is constructed from the following principles: the objects and the processes are the primary entities of the physical world; the mental entities; the mental objects are represented by the communication process via physical objects and actions; the social organization and processes involve roles that are performed by agents that are identified as individual persons. CLO is based on DOLCE+, an extension of the DOLCE foundational ontology [13]. The development of CLO takes into account methodologies proper to foundational ontologies, and proposals in the fields of legal ontologies [10]. CLO organizes juridical concepts and relations on the basis of formal (meta) properties defined in the DOLCE foundational ontology [4]. The basic types of entities that populate the domain of Law are assumed to be clearly identifiable and reasonably intersubjective, and, as such, they are highlighted through a minimal set of properties and relations from DOLCE+. The methodological choices, as well as the exploitation of properties suitable for the legal domain, are based on the approach of legal theories and the philosophy of law.

The development process of these legal ontologies (concept modeling, reasoning in legal domain) is based on legal theories [14], [15]. The presented work does not follow a legal theory. It uses the TIA method [16] based on legal text analysis. The obtained resource and the used method differ from [17]'s work. Indeed, in her work the resulting resource is a linguistic ontology whereas our ontology is formal.

3 The construction method

The ontology construction is based on the TERMINAE method. After selecting a corpus, the method proposes to obtain, at the linguistic level, the linguistic knowledge (terms, lexical relations and groups of synonyms. The linguistic knowledge is recorded in a terminological form. This linguistic knowledge is then transformed into a semantic network during the normalization activity. The link between the text and the concept is reachable through the terminological form. The semantic network includes concepts, relationships between concepts and attributes for the concepts, and is implemented in the formalization activity.

The activities proposed by the TERMINAE method can be split in four steps: constitution corpus, linguistic study, normalization and formalization. Normalization is a particular conceptualization process based on corpus analysis, in line with [18] in contrast with expert introspection. In a second step, the expertise and the target system influence concept definitions. Indeed, the restricted meaning of concepts is mainly derived from the study of term occurrences in texts. These terms become concept labels. Thus, concepts are described thanks to the use of their labels, together with the other terms in the corpus. So, the corpus plays an important role during normalization. Linguistic study and normalization are closely intertwined and unfold in a cyclic way. Throughout the normalization process, linguistic tools and principles are used to explore the text and to decide whether a concept, an attribute or
a relation should be defined or not. The semantic normalization begins with the study of terms described as central. These terms engender central concepts. Normalization involves two development axes, a bottom-up axis which is similar to alignment in merging ontology processes and a top-down axis. The activity performed through the latter axis is to create concepts which specialize and differentiate already defined concepts.

The last step of the method is formalization. The ontology language is included in the TERMINAE tool and is similar to description logic [19]. Concepts are defined by necessary and sufficient conditions, in order to be organized in a subsumption taxonomy and classified in this hierarchy according to their properties. A property, named role, describes a binary relation between a “linking” concept and a “value” concept (called role value). Individual concepts represent instances of concepts and express the value for their roles. Moreover, to support the understanding and maintenance of the modeling, concepts are given specific labels to express whether they are structuring or linguistic. Structuring concepts are introduced to structure the ontology whereas terminological concepts are built from the corpus study, and they correspond to one or several domain terms, one of them being the concept label. These concepts are linked to their term occurrences in the corpus. The TERMINAE tool provides material support to normalization and formalization activities. Moreover, TERMINAE allows importing and exporting ontologies in owl language, which aids the alignment process. SYNTAX [20] is used in the linguistic study activity to obtain domain terms. The LINGUAE concordancer included in TERMINAE allows pattern recognition. MFD [21] uses association rules to find lexical relations. The alignment process takes place during the normalization activity. The mismatches that may exist between separate ontologies [22] enhance the difficulties in merging and aligning. Mismatches may appear at two levels: (1) language and (2) ontology. At the first level, the kinds of mismatches pertain to: syntax, logical representation, semantics of primitive concept and language expressivity. At the second level, mismatches are differences in the way the domain is modeled. The found mismatches pertain to conceptualization, explanation and terminology. Apart from these problems, there are practical problems: it is difficult to find the terms that need to be aligned; the consequences of a specific mapping are difficult to see. [3] defined the set of basic operations for ontology alignment as: removing a class from the ontology; removing a concept from the list of parent concepts because a more appropriate parent exists in another ontology; adding a concept to the list of parents of the considered class because an appropriate parent is found in another ontology; renaming a concept to conform to naming conventions in another ontology; moving a concept from one ontology to another; removing a slot from the list of slots of a concept because a similar slot was inherited from a new parent; moving a slot because it is more appropriately defined in a parent; renaming a slot. Our ontology is built around two central ontologies DOLCE and LRI-Core. Links are established between DOLCE and LRI-Core upper level ontologies. The resulting ontology does not become part of the DOLCE ontology but uses its top-level distinctions. The process of ontology alignment was executed during the ontology construction and was performed mostly by hand, with the TERMINAE interface. TERMINAE provides easy import of concepts in DOLCE but doesn’t check whether consistency is maintained once the operations are performed.

4 First results: description and analysis

4.1 The corpus constitution

The corpus consists in a single text of seven pages selected by jurists, namely the ”Council directive 2001/23/EC of March 12th, 2001 on the approximation of the laws of the Member States related to the safeguard of employees’ rights in the event of transfers of undertakings,
businesses or parts of undertakings or businesses”. It is written by the European Union Council, which is the first decision organ of the Union. The linguistic study is based on the French version but the English version was used to translate our examples.

4.2 Linguistic study

This step consists in selecting the terms and lexical relations that will be modelled. The results of this stage are quite raw and will be further refined. SYNTEx, linguae and mfd are the used tools. SYNTEx yielded 900 candidate-terms. The most used terms were : travailleur (employee) (56 occurrences), transfert (transfer) (41 occurrences), cédant (transferor) (15 occurrences), cessionnaire (transferee) (14 occurrences). Because of a small corpus, there are many hapax (a term which appears only once in the corpus) which are significant and have to be kept such as the syntagm community charter of the Fundamental Social rights or the verb to abrogate. LINGUAE and MFD tools have been used to explore the relations between central concepts. For example, the relation between transferor and transferee is given by a pattern of linguae (see Figure 1).

4.3 Semantic normalization

The aim of the ontology building is to describe the concept of TRAVAILLEUR(EMPLOYEE) in the context of the outsourcing of his/her enterprise in the European community. Our results are based on the above method described. We start with three central concepts (TRAVAILLEUR(EMPLOYEE) ; TRANSFERT(TRANSFER), LICENCIEMENT(RENDUNDANCY)).

a. Modeling bootstrap
From the terminological form of the term bound to a central concept, we find terms and relations describing the concept. Terms are translated into concepts and relations into roles. A definition in pseudo-natural language is established. The concept definition is given, in comprehensive form, from the term occurrences found in the text with structural or functional properties. A structural property references an ancestor concept property. A functional property describes a relation between concepts. The following figures present concept definitions obtained by using the text. The used syntax simplifies the linguistic expression to be closer to the ontology language.

Study of TRAVAILLEUR (EMPLOYEE)

The terminological form involves the definition shown in figure 2. This definition shows three important points : (i) the European community is constituted of member states which have national laws ; (ii) an employee is protected by national employment laws ; (iii) there
Sylvie Despres and Sylvie Szulman

is no explicit link between national laws and European laws. All these elements have to be represented in the ontology. From occurrences of the term *employee* (see figure 3), the definition of the concept *EMPLOYEE* has been established:

<table>
<thead>
<tr>
<th>TRAVAILLEUR</th>
<th>EMPLOYEE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROPRIÉTÉS STRUCTURELLES</strong></td>
<td><strong>STRUCTURAL PROPERTIES</strong></td>
</tr>
<tr>
<td>est une personne</td>
<td>is a person</td>
</tr>
<tr>
<td>appartient à une entreprise</td>
<td>belongs to an enterprise</td>
</tr>
<tr>
<td>appartient à un Etat membre</td>
<td>belongs to a member state</td>
</tr>
<tr>
<td>il existe plusieurs catégories de travailleurs</td>
<td>several kinds of employees exist for legislation</td>
</tr>
<tr>
<td><strong>PROPRIÉTÉS FONCTIONNELLES</strong></td>
<td><strong>FUNCTIONAL PROPERTIES</strong></td>
</tr>
<tr>
<td>possédant des droits et des intérêts</td>
<td>having rights</td>
</tr>
<tr>
<td>ayant des représentants</td>
<td>having representatives</td>
</tr>
<tr>
<td>ayant des conditions de travail</td>
<td>having working conditions</td>
</tr>
<tr>
<td>est l’objet du transfert de son entreprise</td>
<td>is the object of the transfer of his/her enterprise</td>
</tr>
<tr>
<td>est protégé par une législation nationale</td>
<td>is protected by national employment law</td>
</tr>
</tbody>
</table>

**FIG. 3: Concept TRAVAILLEUR(EMPLOYEE) definition**

Study of the term *transfert* (*transfer*)

The definition of the concept *TRANSFER* was obtained in a similar way:

<table>
<thead>
<tr>
<th>TRANSFERT</th>
<th>TRANSFER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROPRIÉTÉS STRUCTURELLES</strong></td>
<td><strong>STRUCTURAL PROPERTIES</strong></td>
</tr>
<tr>
<td>est exécuté à une date donnée et a une durée dans le temps</td>
<td>is executed at a given date and has a duration</td>
</tr>
<tr>
<td>ayant un agent source (cédant) et un agent destination (cessionnaire)</td>
<td>having a source agent (transferor) and a destination agent (transferee)</td>
</tr>
<tr>
<td><strong>PROPRIÉTÉS FONCTIONNELLES</strong></td>
<td><strong>FUNCTIONAL PROPERTIES</strong></td>
</tr>
<tr>
<td>porter sur un objet (entreprise)</td>
<td>focuses on an enterprise is protected by national laws</td>
</tr>
<tr>
<td>est protégé par une législation nationale</td>
<td>is protected by national laws</td>
</tr>
</tbody>
</table>

**FIG. 4: Concept TRANSFERT (TRANSFER) definition**

The link between the terms *transferor* and *transferee* was drawn from the occurrences of the term *transfer*. The definitions of "transferor" and "transferee" appear in the directive (see figure 4).

Study of the term *licenciement* (*redundancy*)

There exists a great proximity between the terms *transfer* and *redundancy*. The concept of *REDUNDANCY* is also linked to the concept of *PROCESS*. The study of this term shows a link with the term *transfer*. The concept of *LICENCIEMENT (REDUNDANCY)* (see figure 5) is also bound to the concept *PROCESS* but differs from it because its object is persistent whereas it is not persistent for the concept *REDUNDANCY*. 
During the activity of modeling consolidation, the alignment process is integrated in the semantic normalization phase. The used ontologies are either a generic ontology as DOLCE and a legal one as LRI-Core. The domain modeling requires the introduction of non terminological concepts for structuring the ontology. The work according to the bottom-up axis is dedicated to specialization and differentiation of the concepts already defined. The roles describe the functional properties. Some of them restrict inherited roles.

Generalization and attachment of the EMPLOYEE concept

The structural property "is a person" of the EMPLOYEE concept leads to studying this concept in the legal domain. The definitions of the terms juridical person and natural person don't appear in the directive but are necessary for constructing the ontology. Recourse to the expert helped us to alleviate this lack of definition. The structural properties: (a) "belong to" of the EMPLOYEE concept leads to the study of the terms enterprise and member states; (b) "there exists several employee categories" was studied according to the top-down axis for identifying these different categories and refining the concept of EMPLOYEE.

Generalization and attachment of the TRANSFER concept

The structural properties of TRANSFER bring out characteristics of the general concept of PROCESS as (agent, object, source, destination, date, duration). The PROCESS concept from DOLCE was added to the ontology as a parent of the TRANSFER concept. This add-parent operation requires the introduction of the concepts of DATE and DURATION that are related to a temporal description of the events. TRANSFEROR and TRANSFEREE which can be a juridical person and a natural person subsume respectively the source agent concept and the source destination concept. These concepts are defined in the LRI-Core ontology.

Introduction of concept and structuring property

To complete the model, the structuring and non terminological concepts of SOCIAL OBJECT and JURIDICAL OBJECT and the terminological properties govern and to be governed by have been introduced.

The concept of SOCIAL OBJECT, represented in the same way in DOLCE and LRI-Core is used to attach the concept PHYSICAL PERSON. The term juridical or natural person has been translated into the concept of PERSONNEPHYSIQUEMORALE. This concept is attached to the concept JURIDICAL OBJECT (OBJETJURIDIQUE). It is not represented in LRI-Core therefore.
we have defined it as a DOCUMENT because a JURIDICAL OBJECT is described by means of texts.

The concept specialization

To complete the ontology, taking into account general knowledge, the hierarchy under the concept REDUNDANCY has been specialized by the terminological concept collective REDUNDANCY and by the terminological-non-attested concept individual REDUNDANCY.

Figure 6 presents an extract of the top of the built ontology and the links with the DOLCE and LRI-Core concepts.

Role definition

The roles govern and to be governed have been created to define the concepts of SOCIAL OBJECT and JURIDICAL OBJECT. During the ontology construction, it appeared that the term govern was used in the directive with the agent JURIDICAL OBJECT (a contract of employment) and with the object SOCIAL OBJECT specialization (employment relationships). Lexical relations as covered by or protect that constitute specializations of the relation govern (see figure 7) were also used in the text. Therefore a role hierarchy was created between "govern" and "protect" roles.
This first version of the ontology includes around one hundred concepts and thirty roles. Figure 8 presents an extract of this ontology.

\[\text{topConcept} \]

\[\text{PERDURANT} \]

\[\text{STATIVE} \]

\[\text{processus} \]

\[\text{transfert} \]

\[\text{restructuration} \]

\[\text{licenciement} \]

\[\text{licenciementIndividuel} \]

\[\text{licenciementCollectif} \]

\[\text{ENDURANT} \]

\[\text{PHYSICAL-ENDURANT} \]

\[\text{PHYSICAL-OBJECT} \]

\[\text{objetJuridique} \]

\[\text{personnePhysiqueMoral} \]

\[\text{cédant} \]

\[\text{cessionnaire} \]

\[\text{paragraphe} \]

\[\text{NON_PHYSICAL_ENDURANT} \]

\[\text{NON_PHYSICAL_OBJECT} \]

\[\text{objetSocial} \]

\[\text{restructuration} \]

\[\text{personnePhysique} \]

\[\text{personneTravailleure} \]

\[\text{travailleur} \]

\[\text{personneResponsable} \]

\[\text{organisation} \]

\[\text{étatMembre} \]

\[\text{représentantDeTravailleure} \]

\[\text{entreprise} \]

\[\text{licenciement} \]

\[\text{licenciementIndividuel} \]

\[\text{licenciementCollectif} \]

\[\text{emploi} \]

\[\text{relationDeTravail} \]

\[\text{contratIntérimaire} \]

\[\text{CDD} \]

\[\text{conditionDeTravail} \]

\[\text{conventionCollective} \]

\[\text{ABSTRACT} \]

\[\text{REGION} \]

\[\text{PHYSICAL_REGION} \]

\[\text{SPACE_REGION} \]

\[\text{communautéEuropéenne} \]

\[\text{FIG. 8: An ontology extract} \]

5 Conclusion

This paper presents the building of a legal ontology about the concept of EMPLOYEE in the European Community. This work was initialized by jurists. They wanted to establish the relationships between European and national legislations around the concept of EMPLOYEE. We have conceptualized from a directive the notion of employee in the restrictive context of enterprise transfer. During the ontology construction, experts didn’t participate actively in the ontology construction and their demand were not very specific but they validated the obtained resources. Moreover, it appears that the questions asked for modeling the concepts are connected with the points examined by jurists (for instance, the link between a TRANSFER and a PROCESS). It is interesting to note that without resorting to any legal theory we
obtained a satisfying model of the domain. As the E-POWER [23] project, the finality of our work is to "translate" legislation into a formal specification that can be used by computers, but our method is semi-automatic. We essentially differ from the work of Querasma team [24] by task object of the ontology building. We followed the construction method from texts (TERMINAE) in which we introduced an alignment process to core ontologies. This alignment process is complementary to the textual approach because it allows us to reuse general concepts and meta-properties defined in the core ontologies. The TERMINAE tool permits a semi-automatic treatment of the alignment process with human intervention.

The alignment process in this case included the following activities: the identification of the content that overlapped with the core ontology; the concepts that were at the top level became subclasses of more general concepts. The concepts are defined from the study and interpretation of the term occurrences in the directive. The term properties (structural and functional) are translated into a restricted language. This translation was realized by hand. The linguistic criteria for identifying these properties remain to define to automate this process.

Now we project to study the directive written in English to compare with the obtained model from the text written in French. The question is whether a single model of the concept EMPLOYEE, independent of the used language, exists or not.

Références


