Towards Semantic Interpretation of Legal Modifications through Deep Syntactic Analysis

Raffaella BRIGHI [a], Leonardo LESMO [b] Alessandro MAZZEI [b],
Monica PALMIRANI [a] and Daniele P. RADICIONI [b],

[a] *Università di Bologna, CIRSFID*
[b] *Università di Torino, Dipartimento di Informatica*

Abstract. We are concerned with the automatic semantic interpretation of legal modificatory provisions. We propose a novel approach which pairs deep syntactic parsing and a fine-grained taxonomy of legal modifications. Although still in a developmental stage, the implemented system can be used to annotate with meta-information modificatory provisions of NormaInRete documents.

Keywords. Legal documents semantics, natural language processing, automatic information extraction.

Introduction

Many efforts have been made in the last few years for investigating two problems in the analysis of legal texts: how to automatically identify structural portions of legal documents as long as their mutual references [4,3]; and how to grasp semantic information [1,9,5,2]. Various national initiatives have established XML standards for describing legal sources and documents, which have grown into projects aiming at integration and interoperability across all legal domains.

Since hand-made annotations are time-consuming and error-prone, tools for modeling and representing the structure and content of norms are needed. Such tools could greatly benefit from automatic approaches to extract both *structural* and *semantic* data from legal texts, conceivably generating XML output. In the Italian legal domain a *regular* structure has been identified in legal provisions, thereby making possible to use a text editor to mark up semi-automatically structural partitions and normative references. Our present research is grounded on the *NormeInRete* standard (NIR), for the Italian Legal Text.[1]

In this paper we single out a subset of the required annotations, namely the annotation of *modificatory provisions*. Since a normative system can be thought of as a body of norms changing over time, when new norms are introduced, existing ones get changed or ejected (see [8] for a theory of *diachronic* normative system). Thus, it will often happen that one norm refers to other norms, either to find its own completion (through the con-

---

[1] http://www.normeinrete.it/.

```
<dsp:sostituzione>
    <dsp:pos xlink:href="#art1-com4" />
    <dsp:norma xlink:href="urn:nir:stato:legge:2005-12-28;262"> <dsp:pos xlink:href="#rif9"/></
    dsp:norma>
    <dsp:novella><dsp:pos xlink:href="#mod16-vir2" /></dsp:novella>
    <dsp:novellando><dsp:pos xlink:href="#mod16-vir1" /></dsp:novellando>
</dsp:sostituzione>
<comma id="art1-com4">
    <num>4.</num>
    <corpo> All'<mod id="mod16"><rif id="rif9" xlink:href="urn:nir:stato:legge:2005-12-28;262#art40-
    com1">articolo 40, comma 1, della legge 28 dicembre 2005, n. 262</rif>, le parole: <virgolette tipo="parola"
    id="mod16-vir1">"sei mesi"</virgolette> sono sostituite dalle seguenti: <virgolette tipo="parola" id="mod16-
    vir2">"dodici mesi"</virgolette></mod>.</corpo>
</comma>
```

Figure 1. A substitution provision with structural and semantic markup for the Italian sentence *All'articolo 40, comma 1, della legge 28 dicembre 2005, n. 262, le parole "sei mesi" sono sostituite dalle seguenti "dodici mesi" (At the article 40, comma 1 of the law December 28, 2005 number 262, the words "six monts" are to be substituted by the following "twelve monts").*

tent referred to) or to change these other norms text, purview, or term of enforceability through a modificatory clause embedded in the norm itself.

Although the legal language is more controlled than ordinary language, tracking modifications requires considering the deep syntactic structure of sentences (see, e.g., [7]). Moreover, it involves encoding prior knowledge on the sorts of possible modifications, and of the manners such modifications can practically occur in the legal discourse. Therefore, the problem of annotating modificatory provisions is still a –relevant– open problem.

The system for the semantic interpretation presented here has the following strengths: it builds on a deep parsing approach (based on the Turin University Parser, TUP [6]) and takes advantage of a fine-grained taxonomy of modificatory provisions. Furthermore, based on the taxonomy of modificatory provisions, we devised some semantic frame for each type of modification (e.g., for *deletion*, *substitution*, etc.). Each semantic frame has associated its own legal semantic roles (*slots*) to be filled: we adopt a modular rule-based approach for filling the slots associated to the appropriate semantic frames.

1. A taxonomy for modificatory provisions

The NormeInRete (NIR) standard includes a part dedicated to modifications. Figure 1 illustrates how a non-qualified provision can be enriched with semantic metadata (in bold) by marking it up in XML through NormeInRete.

Before any semantic annotations are added, the text is marked up with structural data as well as with normative references and "quotation mark" elements, i.e. pieces of text referring to a passive norm (such as any additional or replacement text, along with a string indicating where this text belongs in the passive norm). Semantic annotation enriches a text with the <mod> element, which delimits a modificatory clause and with the metadata, that fully qualify the modification and its attributes according to one of the following classes:

Type 1. A change made to the norm text or form (an *integration*, *replacement*, *deletion*,

relocation) or to the norm meaning (an *interpretation* or *variation of meaning* or a *modification of clauses*);

Type 2. A change made to the range of a norm (an *extension* of its subject matter or range of application or a provision stating a *derogation* to it);

Type 3. A change made to the *temporal parameters* of the norm (the time of its entry into force, and the time when it becomes applicable or effective);

Type 4. A change made to the *status* of the norm within the legal system (a decree-law that is made into law, an international treaty that is transposed into domestic law);

Type 5. A change made to the *powers* conferred under a norm within the legal system (e.g., a EU directive transposed into domestic law).

This classification was compiled by working together the modificatory forms described in legal theory and legal informatics [10], on the one hand, and the schemas functional to the consolidation of normative texts, on the other one.

In addition to carrying out a legal-theoretical analysis, we have also surveyed a large body of norms (about 5000 documents) that legal practitioners have semantically annotated on the basis of the NIR format. Our specific purpose was to investigate the effective usage of legal-theoretical categories, the language used and composition rules. Furthermore, thanks to the regularity of the language used in active modificatory provisions, we have individuated and encoded some idioms that frequently recur in the texts analyzed. The common idioms that recur in the different modificatory provisions have been sorted into different classes of modificatory language, and to each class we have associated one or more legal classes. The standard locutions are often accompanied by other recurrent elements that specify the relevant modificatory action: *Date*, *Quoted text*, *Position*, *Condition*. Such information is exploited by the semantic interpreter, introduced in next Section.

2. Annotating modificatory provisions

We preliminarily retrieve the location of a modificatory provision within the document, and then simplify the input text by replacing text fragments with their corresponding identifiers. For example, by applying this rewriting procedure to the sentence in Figure 1, we obtain the text "`All'RIF9, le parole VIR1 sono sostituite dalle seguenti VIR2`", which will be processed by the syntactic parser.

To annotate modifications with meta-information, we have devised a two-step process. We first perform a syntactic analysis of the retrieved provision; then we semantically annotate the retrieved provision by using its syntactic structure and based on the modificatory provisions taxonomy introduced above.

2.1. Syntactic Analysis of Modificatory Provisions

The TUP is a rule-based parser that returns the syntactic structure of sentences in the dependency format. Dependencies are binary relations (e.g. *subject-relation*) between a dominant word (the *head*, e.g. the verb) and a dominated word (the *dependent*, e.g. the noun-subject). After the *morphological analysis* and the *part of speech tagging* the input sentence undergoes three phases for hanling *chunking*, *coordination* and *verbal subcategorization* [6]. Such a tree dependency representation is the basis for the final step of semantic interpretation.
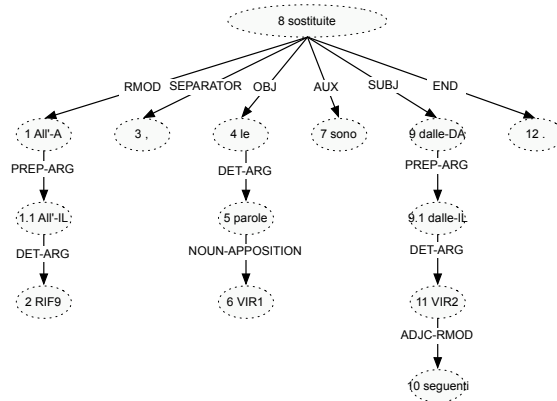
Figure 2. Syntactic analysis of the sentence: *All'RIF9, le parole VIR1 sono sostituite dalle seguenti VIR2 (At the RIF9, the words VIR1 are substituted by the following VIR2).*

2.2. Semantic Interpretation: Filling Semantic Frames

The semantic interpreter checks if the root of the syntactic tree is a verb belonging to the modificatory provisions taxonomy (Section 1). In this case we have a fundamental cue that the sentence being analyzed contains a modificatory provision, and we retrieve the semantic frame associated with the verb. Then the semantic interpreter first uses a set of pattern matching rules on the syntactic tree to recover the relevant parts of the syntactic substructures (e.g., the text fragments references). It then matches these relevant parts with the slots of the semantic frame.

Consider for example the sentence: *All'RIF9, le parole VIR1 sono sostituite dalle seguenti VIR2 (At the RIF9, the words VIR1 are substituted by the following VIR2)* and the corresponding syntactic tree in Figure 2. The root node of the syntactic tree is the verb "sostituire" (*substitute*). As a consequence, the semantic frame associated to the verb class *substitute* is selected: since the semantic roles (i.e. the slots of the frame) are related with the legal category *substitution*, the verbs *substitute*, *change*, *modify*, etc. have the same frame. Three pattern matching rules match the syntactic tree in Figure 2; each rule recovers a relevant part to fill a slot in the semantic frame associated to the verb "sostituire". The first rule fills the semantic role `referenceDocument` with the string "*RIF9*". This rule is activated by the conditions that *i*) the syntactic relation connecting "All' " to the verb is *RMOD* (that is, *modification relation*); and *ii*) that there is a descendent of the preposition which is a reference (*RIF9*). The second rule fills the semantic role `modifiedText` with the string "*VIR1*". This rule is activated by the conditions *i*) that the syntactic relation connecting "il" to the verb is *OBJ* (*object relation*); and *ii*) that there is a descendent of the preposition which is a reference (*VIR1*). The third rule fills the semantic role `modifyingText` with the string "*VIR2*". This rule is activated by the conditions that *i*) the syntactic relation connecting "dalle" to the verb is *SUBJ* (*subject relation*); and *ii*) that there is a descendent of the preposition which is a reference (*VIR2*). No further rules can be applied to the structure, and the semantic interpretation algorithm returns the following semantic frame.

```
FRAME sostituire

legalCategory:        substitution
referenceDocument:  RIF9
modifiedText:         VIR1
modifyingText:        VIR2
```

Some further rules are designed to account for more complex linguistic constructions. For instance, in the case of coordination we need a number of special rules to duplicate the information of the frame.

3. Conclusions and Future Works

In this paper we have presented a novel approach for the automatic annotation of modificatory provisions. We use a deep parser for Italian, and we adopt a rule-based algorithm in order to fill the semantic roles of the semantic frame associated with the modificatory provision. Our approach has some similarities with a number of previous works [2,1,7]. The work in [2] is different from our system in that it only uses shallow syntactic analysis for annotating general provisions; the work in [1] encodes the meaning of modificatory provisions with semantic frames, but it is related to the automatic generation of modificatory provisions rather than with their analysis; finally, the work described in [7] uses a deep syntactic parser but it aims at a full semantic interpretation of judicial opinions.

Our system is still in a prototypal stage; in the next future we will enlarge the set of rules in order to have a wider coverage of syntactic phenomena. Moreover, we plan to perform quantitative evaluation of the system by using a "gold standard" approach, i.e. testing against set of hand-annotated modificatory provisions to quantify the number of correct annotations produced by the system.

References

[1] T. Arnold-Moore. Automatic generation of amendment legislation. In *Procs. of ICAIL*, pages 56–62, 1997.

[2] C. Biagioli, E. Francesconi, A. Passerini, S. Montemagni, and C. Soria. Automatic semantics extraction in law documents. In *Procs. of ICAIL 2005*, pages 133–140, New York, NY, USA, 2005. ACM.

[3] C. Biagioli, E. Francesconi, P. Spinosa, and M. Taddei. The NIR project: Standards and tools for legislative drafting and legal document web publication. In *Procs. of ICAIL Workshop on e-Government*, pages 69–78, 2003.

[4] A. Bolioli, L. Dini, P. Mercatali, and F. Romano. For the Automated Mark-up of Italian Legislative Texts in XML. In T. Bench-Capon, A. Daskalopulu, and R. Winkels, editors, *Procs. of Jurix 2002*. IOS Press, 2002.

[5] E. de Maat and R. Winkels. Categorisation of norms. In A. R. Lodder and L. Mommers, editors, *Jurix 2007: The Twentieth Annual Conference Annual Conference*, pages 79–88. IOS Press, 2007.

[6] L. Lesmo and V. Lombardo. Transformed Subcategorization Frames in Chunk Parsing. In *Proc. 3rd Int. Conf. on Language Resources and Evaluation (LREC 2002)*, pages 512–519, Las Palmas, 2002.

[7] L. T. McCarty. Deep semantic interpretations of legal texts. In *Procs. of ICAIL 2007*, pages 217–224, New York, NY, USA, 2007. ACM.

[8] M. Palmirani and R. Brighi. Time model for managing the dynamic of normative system. *Electronic Government*, pages 207–218, 2006.

[9] J. Saias and P. Quaresma. A Methodology to Create Legal Ontologies in a Logic Programming Based Web Information Retrieval System. *Artificial Intelligence and Law*, 12(4):397–417, 2004.

[10] G. Sartor. *Il procedimento normativo regionale*, chapter Riferimenti normativi e dinamica dei nessi normativi. Cedam, 1996.