Formal Models for a Legislative Grammar. Explicit Text Amendment

Andrea Bolioli[1], Piero Mercatali[2], Francesco Romano[2]

[1] CELI, Language & Information Technology, Corso Moncalieri 21,
10131 Torino, Italy
bolioli@celi.it
[2] ITTIG CNR, Istituto di Teoria e Tecniche dell'Informazione Giuridica del Consiglio Nazionale delle Ricerche, Via Panciatichi 56/16,
50127 Firenze, Italy
{mercatali, romano}@ittig.cnr.it

Abstract. In this paper we will present a method for mining information within legal texts, in particular in regards to corpora of statutes. Text mining, or more in general Information Extraction, can provide a valuable help to people involved in research about the linguistic structure of statutes, and, as a side effect can be the seed for a new generation of applications for the validation and conversion in the legislative domain.

1 Scope and Assumptions

For the communication of legislative sources through the Internet, the parliamentary and governmental institutions of many countries[1] have begun a process of converting their "deposits" of these into a standard format for facilitating the retrieval and display of texts.

The XML mark-up language seems to be the tool deputised for reaching this scope[2]. In fact, this language combining its dual nature as a mark-up language and a Web standard, is able to form the common ground for action both "at the source", namely, legislative drafting, and action "downstream" relating to the publication of the texts and the identification of tools for accessing legislative information [2] [10]

In Italy, the introduction of the XLM language for processing legislative instruments was proposed and experimented in the "Norme in rete" [Law on the Net] Project, solicited by the

[1] See, for example: the data bank of legislative instruments of the Australian State of Tasmania at http://www.thelaw.tas.gov.au and the DTD relating to the parliamentary acts in the United States on the site http://xml.house.gov.

[2] Although standard languages are available, it is nevertheless necessary to reconcile the use of these languages for the man-machine dialogue with the citizen's access to and fruition of information. Access to and fruition of the mass of public data, although transmitted by machines, cannot be achieved with artificial languages, but only with non specialised communicative formats like natural language.
Therefore , it appears that the automated processing and recognition of natural language, has a fundamental role to play in this man-machine interaction.

Ministry of Justice, financed by AIPA (Agenzia per l'Informatica nella Pubblica Amministrazione) [Agency for Informatics in the Public Administration] and developed under the guidance of the Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) of the Italian National Research Council. The project has produced, amongst other things, the DTD rules adopted as a standard by AIPA3 for the on-line publication of Italian legislative instruments.

In order to adopt this language as a standard and, above all, for the conversion of the legislative instruments in force into the format provided for by the DTD rules, two factors, in our opinion, must interact.

Definition and promotion of a "controlled" legislative language

Rules for law-making or techniques for legislative drafting have introduced unambiguous and recurrent elements into legislative instruments, whereby it is possible to identify a more controlled language in legislative language compared to natural language. In fact, specific rules of orthography, lexicon, syntax, style and structure for the drafting of legislative instruments have been adopted. A collection of these rules is to be found in a Circular[4] issued, in 2001, by the President of the Council of Ministers, and the Speakers of the Chamber of Deputies and the Senate and adopted by the Government and Parliament. The Circular updates an earlier one of 1986[5]. For drafting their legislative measures, almost all the Regions in Italy have adopted the "Rules and Suggestions for Drafting Legislative Texts" Manual, a set of rules that are almost the same as the state rules, compiled in 1991 and updated in 2002.

These rules have been applied and complied with in the drafting of legislative instruments enacted by the State and Regions since the end of the 1980's[6]. Leafing through the legislative documents, it cannot be said that these rules have, up until now, been strictly and uniformly applied by all law-makers. However, some analyses of sample texts have shown that the use of the legislative drafting rules is spreading.

The drafting of other legislative documents (such as the regulations of local authorities, collective contracts, etc.) is not bound by these rules. It can, however, be said that it is widespread, in practice, to make reference to these drafting rules, even if their application depends on the sensitivity and knowledge of the draftsman. On the other hand, many initiatives are underway for the formal and binding adoption of the State/Regional drafting rules by all those persons who produce legislative documents. The "Norme in rete" [Law on the Net] Project has contributed to accelerating the process for spreading and receiving the

---

[3] AIPA Circular, 22 April 2002, "Formato per la rappresentazione elettronica dei provvedimenti normativi tramite il linguaggio di marcatura XML" [Format for the Electronic Representation of Legislative Provisions by Means of the XLM Mark-up Language]. The text can be consulted at : http://www.normeinrete.it/standard/Circular-xml.htm.

[4] Circular 20 April 2001, no. 10888 of the Presidency of the Council of Ministers, "Regole e raccomandazioni per la formulazione tecnica dei testi legislativi" [Rules and Recommendations the Technical Formulation of Legislative instruments], published in the Gazette Ufficiale No. 97 of 27 April 2001. The same rules have also be adopted by the Chamber of Deputies and the Senate with identical circulars by their relative Speakers.

[5] Circulars of the Speaker of the Senate, the Speaker of the Chamber of Deputies and the Presidency of the Council of Ministers of 24 February 1986 (G.U. No. 123 of 29 May 1986, Supplemento ordinario No. 40). For an in-depth illustration of the rules for legislative drafting in Italy and in Europe see: [16] [17].

[6] For the historical framework of Italian legislative drafting, see [25].

legislative drafting standards, drawing attention to their utility for electronic processing, whilst still maintaining that the main purpose of these rules is to guarantee greater clarity in and ability to understand legislative instruments.

At the same time, research into legal theory, legal language, and legal artificial intelligence have contributed to the definition of the syntactical and semantic structures and to morphological and lexical behaviour peculiar to the legal discourse.

Use of tools for natural language recognition

It is evident that the presence of common rules consolidate the definition of text models, which the interdisciplinary studies we have just mentioned describe with ever increasing exactitude. It is also evident that this modelling assists in the automated recognition of the structures of legislative instruments and their tagging according to the XLM standard. In fact, this tagging will be difficult to obtain from the law-maker as it is extraneous to the tasks and objectives involved in his/her normal activities. If other professionals do it later, it may provoke an often unsustainable increase in the time needed and the costs involved in building and managing the legislative knowledge base, structured according to XLM standards.

It is within the perspective of the implementation of a parsing system efficient for the automated recognition of the structures of legislative instruments and the subsequent tagging and conversion of these texts in XML format that we shall now begin the description of the research presented here.

2. The Method

For this research, the methodological approach can be subdivided into the following phases:

1. identification of the technical tool for the implementation of the parser and the extraction of the information;
2. identification and description of the models, on the basis of legal rules;
3. identification and description of the textual structures expressing the defined legal models;
4. choice of the sample of legislative instruments to be analysed, compilation of the grammar according to the syntax of the pre-selected parser and the automated analysis of the sample.

In the following paragraphs, we shall attempt to define and describe these phases.

2.1 Identification of the Technical Tool for the Implementation of the Parser and for Information Extraction

The suitable tool for the recognition and tagging of a legislative instrument has been identified in the Sophia 2.1 system of parsing which uses the methodology applied to finite state automata and the finite state transducer: this is software which is flexible and configurable and which enables rules and specific models (already defined or in the course of definition) to be formalised.

In particular, we are working with this software on analysing and tagging the first sample of legislative instruments in the following phases:

- normalisation of the entry text, properly tagging all those structures and textual segments that can be recognised on the basis of characters or, in other words, without resort to or consultation of the lexicon-dictionary;
- lexical (syntactical category) and morphological (flexion passages) analysis of the text in input;
- disambiguation of the syntactical category of the words (*Part of Speech Tagging*);
- partial syntactical analysis (called *chunking*), aimed at identifying the minimum syntactical groups present in the text in input and at grouping them in constituents;
- semantic analysis and identification of the relevant conceptual structures in the text in input;
- conversion of the analysed document from the original format (Microsoft® Word, HTML, RTF, txt, etc.) into the XML format, according to the established DTD.

2.2 Identification and Description of the Models on the Basis of Legal Rules

The application of a device for parsing, like the one described here, requires a set of rules to be written for the identification, in the texts, of linguistic structures which are bearers of the information we wish to extract. We can call this the compilation of a specific grammar of the domain or of the corpus of the texts to be analysed.

The grammar is made up of a set of models defining the linguistic structures; in turn, the models include one or more rules representing a linguistic structure which are subsequently compiled according to the syntax of the parser for the text analysis and the information extraction.

Computational linguistics indicates the extraction of the set of rules and models from the corpus of the texts directly linked to the domain to be dealt with as the most efficient method for constructing the grammar [9]. In other words, we are trying to reconstruct rules and models *a posteriori*, extracting examples of linguistic structures.

The application of this methodology to legislative instruments, however, overlooks the specificity of the nature and function of these texts.

The legislative instrument has, by definition, a prescriptive function, or in other words, it influences the behaviour and status of the addressee, who cannot escape from this function.

In virtue of this, the request (it is based on the same principle of representative democracy which legitimates and, at the same time, binds those who draft legislative acts, namely, the legislator) is that the legislative instrument responds to a set of rules that dominate and, at the same time, stand beside, integrate, and sometimes modify the rules that make up natural language (in our case Italian).

These rules can be defined as legal rules, The category is broad and varied; it includes within it rules with prescriptive force, that vary strongly based on the source from which they come, to whom they are the addressees, the sanctions they bring with them, etc.. For example, as we shall see, it is the Constitution (that has the highest rank among the legislative rules) which determines that the legislative delegation

must contain a period of time[7] within which the Government must enact the legislation which is the object of the delegation.

A "list" of the prescriptive nature of these rules can be drawn up, starting from the legislative sources which, in turn, have different degrees of binding power in accordance with that which is defined, more precisely, as the hierarchy of sources[8].

With regard to other rules, the discussion is whether they have or do not have general legislative value. Amongst these, there are Ministerial Circulars and Circulars of other public authorities (for example, the Speakers of the Houses of Parliament). They are, nevertheless, without doubt, legal rules and have prescriptive force with regard to the employees of that Ministry or public body, as they are, in any case, administrative acts. The rules for drafting legislative instruments, which are very important here due to the fact that they have the specific role of regulating the structure of the legislative instrument, are usually enacted with this type of act.

Concerning the validity and effectiveness of the models constructed on the basis of these rules, we would like to add several considerations. We are able to list three characteristics on which their efficiency largely depends:

- *Flexibility* of the model, which must adapt to the many structural, functional and thematic characteristics of legislative instruments, which use that extremely changeable and unpredictable vehicle known as language.
- *Precision* in the definition of the model itself, thanks to the presence of the legal rules which are in many cases detailed and precise and for the prescriptive nature of the legislative instrument, it goes without saying that reconciling flexibility and precision constitutes a crucial point in the construction of the models.
- *Authoritativeness* with which the model must be endowed, in order to be shared by and held to be valid by all the users of the system. It is clear that this authoritativeness is more guaranteed if the model is derived from the legislative rules, in the strict sense, according to their hierarchy whilst it will be less guaranteed by those rules which only the drafter must follow. It could be more or less guaranteed, to a different extent, by models based on consolidated practices, legislative theories shared by the major part of legal authority, and so on. We can, nevertheless, say that greater precision could correspond to "more authoritative" models, greater flexibility should correspond to "less authoritative" models.

---

[7] Article 76 of the Constitution states: "The exercise of the legislative function may not be delegated to the Government unless principles and guiding criteria have been determined and solely for a limited period of time and for defined objects."

[8] Traditional legal authority classifies legislative sources into primary, sub-primary and secondary sources. This distinction then opens the way to further specifications, still the reason for discussion today among jurists who present exceptions and specific cases, making the hierarchy complex and articulated and not always very clear. Without taking into account that this scheme is then subject, over time, to changes in relation to modifications in the powers of the law-makers and to the creation of new law-makers or the abolition of existing ones (in fact, the rarest case). Primary sources include, in order: the Constitution, constitutional laws and regional statutes, ordinary laws and *decree-law*s; regional laws are considered sub-primary sources; within secondary sources, there are government regulations, decrees of the President of the Republic and of the Government (President of the Council of Ministers and Ministers), whenever they have legislative value. For an in-depth and better description of the scheme we have explained briefly here, see: [18] [11].

It should further be noted that very authoritative models have greater communicative value (they are directed at and basically are accepted by everyone), whilst as authoritativeness gradually diminishes, the models assume an always increasingly interpretative value, that is, they are a "subjective" definition of the function of the single textual structures or of the entire legislative instrument.

If we can, therefore, talk about "Communicative Models" and "Interpretative Models", it becomes fundamental to evaluate the how they are used to chose whether to refer to the former or the latter.

For example, for the description of a legislative instrument, for the purpose of communication *erga omnes*, we cannot fail to refer to a "communicative model" and the use of an "interpretative model" could be dangerous and misleading.

2.3 Identification and Description of the Textual Structures Corresponding to the Defined Legal Models

For the implementation of the grammar that will then be utilised by the parser, it is necessary to integrate the models extracted from the legal rules with the linguistic rules. It must, however, be clarified that some linguistic rules, in as far as they regulate the drafting of a legislative instrument, are already received as legal rules, in particular, as law-making rules and, therefore, their binding power is reinforced.

We can say, for example, that the legal rules prescribe that a type of amending provision, is manifested through the action of substituting parts of the text; the linguistic rule of synonymy enables us to say that the action of substitution is expressed through verbs: to substitute, to change, to alter, etc... In this case, the linguistic rule goes to integrate itself with the legal rules, in the construction of an efficient model for the purpose of the function of parsing the text. The integration may, however, also concern cases in which the legal rule is not so much integrated but the grammar of the legislative instrument is completed, going to describe the linguistic structures, to which the legal rules do not correspond[9].

From textual practices[10], we can extract rules that go to form other models of linguistic structures found in legislative instruments (or to integrate those models obtained with the legal rules). These are models extracted on exemplifying bases, resorting to the analysis of texts according to the methodology practised by computational linguistics.

Clearly, these models do not have the same precision and, above all, the same prescriptive force of those constructed on the basis of legal rules. We have, therefore, only used and implemented them in the parser as a residual category.

---

[9] For example, for the *novella,* the legislative drafting rules provide for well identified functions (integration, repeal, substitution) to each of which one or more linguistic structures correspond which are also carefully described in the law-making rules. We can also find *novelle* with a replacing function in legislative instruments: they arrange the repositioning of a part of the text from one point to another of the article. The replacing *novella* is not, however provided for or regulated by the rules on law-making.

[10] The term *practices* is used here in accordance with its common meaning of recurrent behaviour (in our case, the behaviour of recurrent drafting) and not in the technical legal sense that classifies them among the sources of the law.

We call these last models "malformed" models. This is not so much because they are not correct from the linguistic point of view, but to contrast them with those corresponding to the legal rules, indicated in the previous paragraph, which we define as "well-formed".

Nevertheless, legislative instruments contain linguistic structures that do not correspond to the models described, because the texts may contain actual legal and linguistic errors.

In these cases, we complete the grammar to be implemented in the parser, with rules (we define them as "case-based" rules) which represent exceptions to the models: they are actual errors or exceptions to the rule, just like exceptions in linguistic grammars.

2.4 Compilation of the Grammar and the Automated Analyses of the Sample

The compilation of the grammar in the syntax of the chosen parser takes place by using the Workbench of the system described in paragraph 2.1., through drafting legislative rules, that formalise the defined models and permit the automated identification of the described linguistic structures and the information extraction.

The choice of the sample of legislative instruments to be analysed must obviously respond to the representative criteria of the legislative linguistic domain, from which we intend to extract the information. In describing our research, we indicate the following criteria for identifying the sample which is the object of this initial analysis.

The analysis of the sample can then be carried out in subsequent phases in order to:

- evaluate the results obtained;
- integrate and modify the well-formed models defined *a priori*;
- identify and formalise case-based malformed models and rules;
- extend the analysis to a gradually widening *corpus* to verify the efficiency of the parsing system.

3 Initial Analysis and Applications

We have decided to experiment the method described for the automated recognition and extraction of three typical structures of legislative instruments, structures representing:

- legislative delegation;
- express textual amendment or *novella*;
- express external textual reference.

Among the many structures making up a legislative instrument that are defined by other legal rules[11], the choice fell on the structures listed above, mainly for two reasons:

1. They are precisely defined structures, both legally and linguistically, by the law-making and other very binding legislative rules.
2. They perform the principal function[12] of legislative links, that is, links between different legislative instruments, a very important function for both the purpose of any action or use of the texts (we are thinking about the compilation of the co-ordinated text here), and for the purpose of the reconstruction of the inter-textual dimension, which, together with the contextual dimension is fundamental in a linguistic textual analysis. It is our intention, in fact, to continue our research, attempting to analyse and formalise the other legislative structures that express links, such as prorogation or suspension.

In the following sections, we shall, therefore, present:

- for legislative delegation, the legal analysis for constructing the model;
- for the *"novella"*, the modelling of the text structure on the basis of legislative and linguistic rules and an initial formalisation of these rules in the parser's syntax.

4 Legislative Delegation

Legislative delegation from Parliament to the Government is provided for in the Constitution (article 76), which lays down the rules, to which the delegation provision must comply:

- it must be conferred on the Government;
- it must contain a term within which the Government has to enact the delegated act;
- it must specify the object of the delegation;
- it must contain the guiding principles and criteria to which the Government has to adhere in the exercise of that delegation.

Parliament can lay down conditions that the Government must comply with (request for opinions, hearings, enactment of other acts, etc..). These conditions are not set out expressly in the Constitution, but legal authority and case law hold that Parliament is free to insert them in the delegation.

It is clear that these rules, having been dictated by the Constitution, have the greatest binding force. A legislative instrument that fails to comply with them could

[11] We have in mind, for example, penalties, financial provisions, and provisions regulating the entry into force and the being in force of legislative acts, etc..

[12] It cannot be said that this function is unique, but it is certainly the most important, apart from being easily identifiable and formalised. For example, legislative delegation has the role of transferring powers from a party to another and also carries out, in support of the former, the connecting role between the delegant act and the delegated act.

be challenged before the Constitutional Court and could be changed only with an amendment of the Constitution and not of an ordinary law or other legislative acts.

An ordinary law (Law of 23 August 1988, No. 400) then specifies the constitutional rules and adds the ways in which some special types of delegations must be complied with. In particular, Article 14 (3) and (4) state:

3. Whenever the legislative delegation refers to many objects open to being regulated separately, the Government may exercise it through more than one subsequent act on the aforesaid objects. In relation to the final term laid down by the delegation law, the Government shall periodically inform Parliament on the criteria it is following in organizing the exercise of the delegation.
4. In any event, whenever the term provided for the exercise of the delegation is more than two years, the Government shall ask the opinion of Parliament on the schemata of the delegated decrees. The opinion shall be expressed, within sixty days, by the Permanent Commissions of the two Houses of Parliament which is competent in the matter, specifically stating any provisions held not to correspond to the guidelines of the delegation law. The Government, within the following thirty days, having examined the opinion, shall re-transmit the texts, with its comments and with any amendments, to the Commissions for a final opinion that shall be given within thirty days.

These provisions add at least three rules to the regulation of the structure of the delegation provisions:

- the delegation may contain more than one separate object;
- the Government may enact more than one delegated act when there is more than one object;
- whenever the term provided for the exercise of the delegation is more than two years, the Government must seek the opinion of the two Houses of Parliament according to a defined procedure.

Furthermore Article, 14 (1) of the same Act lays down that the act (or acts) with which the Government exercises the delegation shall be called a "legislative decree".

The binding force of these rules is no longer that of the Constitution, but of an ordinary law. Therefore, it can be amended by another ordinary law. A court may decide not to apply, for the case it is called upon to decide, a delegation provision finding it to be in conflict with what has been established by Law 400/1988, but the Constitutional Court could not find a provision of delegation not complying with the provisions of Law 400/1988 unconstitutional (and therefore completely void).

It should be remembered that the rule used is a rule of 1988: it does not, therefore, count for prior delegation provisions, just as the constitutional rules for delegations prior to 1948 do not count. The definition of delegation structures before these dates will, therefore, be less binding or they will have to draw on other rules[13].

[13] In the pre-Republic legal order in Italy, the exercise of delegation and the relative delegated acts were regulated by Law 100/1926, even though the institute of legislative delegation was pre-existing and, in general, exists in all legal orders that provide for the separation of legislative and executive powers.

As we have already mentioned, the rules for drafting legislative instruments for Parliament and the Government are to be found in Circular of 20 April 2001, No. 10888 . In particular, Article 2 also deals with legislative delegation and regulates, in detail, the structure of this provision:

"Provisions containing legislative delegations, pursuant to Article 76 of the Constitution shall list the following elements: 1) the addressee of the delegation (the Government); 2) the term for the exercise of the delegation and any term for the enactment of additional or corrective provisions; 3) the object of the delegation; 4) the principles and guiding criteria (that must be separate from the object of the delegation). The term "delegation" shall only be used when there is a legislative delegation with the formula: "The Government is delegated to adopt…". Furthermore, the proper name of the act (legislative decree) to be enacted shall always be given and it shall be specified whether the delegation can be exercised with one or more acts. The delegation provisions shall be found in a special article. An article shall not contain more than one delegation provision".

We have already mentioned the legislative value and prescriptive force of a Circular is, without doubt, of a lesser degree than those of the Constitution or of an ordinary law.

It is, nevertheless, equally as evident that the laws are drafted (or corrected) by the drafter, who, based on the responsibility of his/her job, knows and complies with legislative drafting rules.

It is worth making the observation once again, relating to time, which we already made in relation to the Constitution and Law 400/1988; the Circular is actually of 2001 and, therefore, the legal and linguistic constraints imposed by it on the structure of the delegation can only be found in the most recent legislative instruments and, gradually, as we go back in time, we may come across variable structures.

Keeping in mind the prescriptive and temporal restrictions, to which the set of rules we have attempted to identify and describe are subject, we can argue that they form a well-formed model, from the legal point of view, of the structure of the legislative delegation.

On the basis of this model, we can try to describe the text by inserting the tags or qualifiers of the elements making up the structure, as in the following example:

<ADDRESSEE> The Government of the Republic <ADDRESSEE> <ACTION OF DELEGATION> is delegated to enact, <ACTION OF DELEGATION> <TERM> within eighteen months from the date on which this Law comes into force <TERM>, <DELEGATED/ ACT/S> one or more legislative decrees <DELEGATED/ ACT/S > <OBJECT DELEGATION> laying down additional provisions of the legislation on privacy and personal data protection, <OBJECT DELEGATION> <GUIDING CRITERIA> complying with the following principles and guiding criteria: a) to specify the way in which personal data used for historical, research and statistical purposes shall be processed, taking into account the principles found in ... <GUIDING CRITERIA> (Law 31 December 1996, No. 676).

Such a description should subsequently enable the structures to be identified on the basis of the legal rules to be specified and set out in detail, also from the linguistic [14] viewpoint, and then to move on to the compilation of the rules for the implementation of the parser and the analysis of the corpus, as we shall see in the following paragraphs, for the other structures which are the object of our research.

5 Express Text Amendment or Novella

5.1 Definition of the Structure and the Constituent Elements

Amendment provisions, according to Sartor, fall within the main types of legislative links, classified on the basis of their impact on the legal provision involved. Amendments distinguished from the other large branch of referrals or references, are legislative links characterised by the fact that the active provision affects the passive provision, eliminating it, changing the text or changing the legal significance (whilst leaving the text unchanged). This effect is, instead, lacking in the referral, where the active provision avails itself of the passive provision to complete its meaning, without influencing the latter [23].

In relation to the nature of the impact of the amendment of the provision on the passive provision, we distinguish between textual amendments, time-based amendments (that influence the period of time of the applicability of the passive provision), material amendments, (that amend the legislative content of the passive provision without affecting the text). We shall only look at the first type, the express amendments of the text which, traditionally, lawyers in Italy call the *novelle*.

Indeed, it is perhaps more correct to say that the function of express legislative amendment is expressed through the following three aspects:

- *the structure of the novella, made up of an introductory part*, called subparagraph[15] and a part that contains the express textual amendment;.
- *the characteristics of the amending legislative act* and the amended act: indispensable for subsequently being able to reconstruct the amending links between the different legislative sources;

[14] For example, we shall determine and describe the linguistic expressions through which the term is expressed, which may vary within the flexibility permitted by the legal rule. Therefore, we may have expressions of the kind: "within 6 months of the entry into force of the law"; “within 3 months of the end of the first biennial from when the new law comes into effect” and many others. Nevertheless, we may have other legal rules that help in defining these expressions. For example, it is once again a rule of law-making that requires that the period of time is expressed in months when fixing a term (another structure whose formalisation will also serve in this case).

[15] Understood as the ‘part of the provision that introduces the amendment’: it contains the purview aimed at specifying the relationship (substitution or integration or abrogation) between the provision in force previously and that provided by the textual amendment. The new sub-section generally ends with a colon, followed by the textual amendment placed between inverted commas.

- *the citation* with which the document to be modified is cited, that expresses the legislative reference (also a textual reference), a fundamental element of the amending provisions.

On the basis of the three aspects mentioned here, we have endeavoured to define and describe the qualifying elements of the amendment provision. The description, which is set out here, is derived from the rules for legislative drafting and from the analysis of a sample of approximately 100 amending provisions found in 8 State legislative instruments (the four so-called Bassanini Laws and other legislative instruments related to them), enacted between 1968 and 1999.

Type of amending act: indicates the type (Law, Decree-law, Decree of the President of the Republic, Legislative Decree, etc.) of the legislative act in which the amendment is found. It serves to quickly reconstruct the links between the provisions when there are amendments and it is composed of:

- *Name of the act, Date, Number:* these indicate the essential elements of the amending legislative act, both in the full citation and in its simplified form;
- *Position of the novella*: this is the position, within the amending text, where the amendment provision is found, in order to identify the amendment formula with precision, and also to immediately highlight at what level of the structure it is present.
- *Object of the amendment:* this indicates the object of the amendment in the strict sense (or, in other words, whether the amendment affects the entire act, or a part of it, which paragraph, subparagraph, etc.). This element is also important from the point of view of the structure, because when a part is modified, the effect of the amendment also reflects on that of a directly superior level, in particular, on additions or repeals.

Type act to be amended, composed of:

- *Name, Date, Number of the act to be amended:* these elements indicate the characteristics of the legislative act to be amended, the type of act and the essential elements of the document.
- *Action*: this element describes the action of amendment; it should only take on standard values, sometimes in combination: *repeal, substitution, insertion, addition*, but may take on other values (for example: *replacement*).
- *Expression:* it is the linguistic form with which the amendment is provided for, enclosed by inverted commas or other orthographic signs (colon, brackets, etc.), which delimit the amendments. The expression contains the enunciation that provides for the action, up until the colon that introduces the new text.
- *The text of the amendment,* which, on the basis of the drafting rules, is enclosed within inverted commas and proceeded by a colon.

Furthermore, some textual elements (prepositions, adverbs, conjunctions, etc.) have been identified, which act as connectors and qualifiers of the various elements of the amendment provision [8].

Once the elements making up the amending provision had been identified and described, we were able to propose a classification based on two of the elements we believed to be particularly important: the action of amending and its object.

In particular, on the basis of the action of amending, a distinction can be made among the following: repeal, integration and substitution.

As far as the object is concerned, the amendment, instead, operates on either a part (supra-part, article, paragraph, etc.) or on a part of the legislative discourse. It is obvious that each of the identified actions can operate on both the object *"part"*, and on the object *"part of the discourse"*.

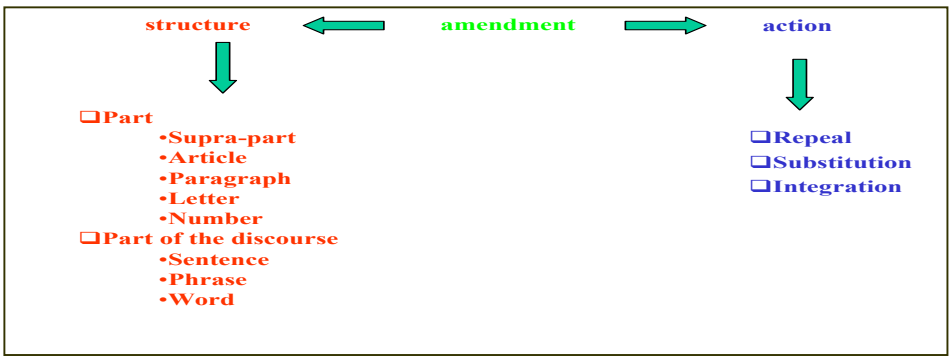

Fig. 1. Definition and structural and semantic classification of the explicit text amending provision

5.2 Formalisation of the Rules and Implementation of the Parser

Each of the types of amendments identified, on the basis of the given classification, was formalised in the syntax of the parser we used. In this way, we got a set of recognition and extraction rules of the "well-formed" amending provisions, based on the models extracted from the Italian regional rules on legislative drafting.

As we mentioned earlier in the introductory part (paragraph 2.3), once the "well-formed" models of amendment were implemented, we moved on to an analysis of a corpus of legislative texts aiming at two goals:

- to verify the validity and flexibility of the formalised rules for the recognition and extraction of the amending provisions, also in cases where the linguistic structure used in the text is not exactly the same as the given model;
- to identify the structures of amending provisions which, although they are logically, legally and linguistically correct, cannot be reduced to formalised models.

For example, the action of integration may, in turn, be divided into actions of addition or insertion. We, therefore, evaluate, on the basis of surveying the corpus, whether to divide the model of the action of integration into two sub-models and, as a result, to implement new rules in the parser.

And, furthermore, during the preliminary analysis of the texts, we found, even if sporadically, the action of relocation which, will probably be inserted in our

taxonomy as an autonomous category and will require us to write a new rule. In fact, although it is not foreseen by the drafting rules, this action seems to perform a specific function of legal amendment and one that is logically correct and expressed with its own linguistic structure.

In writing the rules which implemented the amendment model we mainly used three of the eight modules included in Sophia 2.1 workbench: the compounder, the lexical semantics module and the sentence level semantics module.

In the compounder module, we defined the nominal syntagms important for the purpose of identifying, within the part, the position in which the amendment will act (the final, the last, at the end, before, etc.).

In the second module (LexSem), we, instead, defined the verbal voices, with the various synonyms, corresponding to the various actions of amending (to substitute, to repeal, to insert, to add, etc.).

But it is in the semantics module where it is possible to write the necessary rules for the semantic analysis and for the identification of the significant conceptual structures in the input text.

In fact, in this module, it is possible to build the rules that make up the models to be extracted. The pattern of this module will be made up of the previously assigned semantic categories. These semantic categories also interact with lexical, morphological and syntactic categories, just as with previously defined macros within the semantic module.

Naturally, a variable must correspond to every value attributed to the semantic category which is indispensable in the case where we wish to extract an important datum and therefore to create a template or also where we want to identify the beginning and end of the XML tagging, as in the case of the amendments.

For example the amending structure indicated in:

To Article 4 of Law No, 41 of 28 February 1986, the following paragraph was added, at the end, "6 bis (text of the paragraph)".

is translated into the following pattern of the Sophia parser:

```
REFB+[M_ALLCAT]*+REFE+PUNCTX?+INTEGRATION:startpos+(PUN
CTX+END:endpos+PUNCTX)?+PARAGRAPH_I:vtypenov+PUNCTX?+IN
VERTED COMMAS:startnov+[M_ALLCAT]*+INVERTED
COMMAS:endnov
```

As can easily be seen, the action of amendment was formalised in the semantic category INTEGRATION whilst the syntagm *"at the end"* which indicates the position of the integration of the *novella*, was translated into the semantic category END.

Apart from the category PUNCTX which identifies the punctuation found in the discourse, the other categories identify, in a fairly intuitive way, the inverted commas that delimit the *novella* and the reference to the paragraph, whilst the macro [M_ALLCAT]* identifies all that which is found within the inverted commas.

Two tags have also been introduced which identify the beginning and end of the reference (REFB and REFE). As we have already mentioned (parag. 5.1), the fundamental element of the amending provisions is the explicit legislative text reference with which the document to be amended is referred. The rules of

recognition and extraction of the explicit text references have been implemented in a prior phase of the Project[16], and have been used to define the methodology we have already described and which we are currently perfecting and applying in the formalisation of the amending provisions. These tags are, therefore, only used to identify that portion of the text which contains the reference which is recognised and extracted by a different set of rules.

We intend in this way to obtain a modular formalisation of the text structures, making it possible for every module to integrate, in the phase of the recognition and extraction of the information, with the already created modules with the objective of building an actual grammar that will enable a large number of segments and gradually more and more segments of the legislative discourse to be recognised.

An initial analysis of the text in input will, therefore, allow us to identify the reference found in the amendment and to tag the beginning and end, whilst a second analysis of the same text, will identify the structure to be amended. As already noted, the semantic categories REFB and REFE will enable the parser to identify the beginning and end of the reference found in the amendment.

Therefore, the working hypothesis we are proposing provides for the formalisation of other legislative structures in the syntax of the parser and for integrating the various modules obtained, according to the system we have just described for the amendment and the reference.

In extending this methodology to other parts of the legislative discourse we plan to proceed in such a way as to permit immediate applications which leave aside the construction of a "universal legislative grammar" that, obviously, can only be a long-term objective. For example, the automated formalisation and recognition of references has led to a useful application for building automated links among legislative measures stored in a database or simply available on the Internet[21].

The formalisation of the amending and delegating provisions may lead, within a short time, to other applications, as we shall discuss in the following paragraph.

6 Applications and Future Developments of the Project.

As is well-known, the analysis, recognition and relative tagging of the amending provision are indispensable for the compilation of a co-ordinated text.

The condition to which the repeal exemplified below is subjected, must in some way be picked up by the person who wishes to process the text for co-ordination purposes.

"No. 3) of letter a) of paragraph 1 of Article 5, with regard to competence in the tourism sector, of Law No. 41 of 28 February 1986, is repealed ".

[16] For the description of this phase of the project, see: [5]. It is worth noting here that the initial analyses carried out on laws passed in the 1990s making up part of the selected legislative corpus, confirm, for now, that, thanks to the models of regular citations compiled in accordance with the parser's grammar it was possible to identify and extract over 95% of the explicit legislative text references, conforming to legislative drafting rules.

In this example the repeal is partial and change the meaning of the referred disposition but leaves unchanged the text.

The approach we propose, making use of a linguistic parsing tool, should be able to pick up or at least highlight cases of this kind.

Similar considerations can be made for delegation provisions.

The importance attributed to the processing of this kind of provision, is due to the interest we have in being able to monitor the moment in which the different delegations attributed to the executive will expire[12].

However, the determination of this term creates some difficulties deriving from the way in which the *dies a quem* is expressed. Also, for the purpose of this investigation, it may be thought that it is necessary to resort to tools which allow not only for a purely statistical and probabilistic analysis.

Finally, we believe that we can also apply the methodology we have illustrated here to projects for the control of the quality of legislation.

In fact, we have seen, from several parts, renewed interest in legislative-drafting analysis (LDA) whose aim is to evaluate the quality of the legislative text and its effect on the legislative order in force.

Today, qualitative-type analyses are an important sector of legislative drafting whilst, attempts at LDA that use quantitative methods based on drafting rules and the co-ordination of legislative texts and tending to evaluate the quantity of the errors with respect to the rules are, instead, limited.

These attempts at quantitative evaluation are, however, developing within the technical support structures of legislative institutions and have given encouraging results like those obtained by the Working Group of the Regional Council of Tuscany [17] which has prepared an index of the quality of the regional laws of Tuscany[18].

The Working Group has defined the concept of the quality of laws starting from the assumption that quality is to be understood as the relationship between the text of the law and the legislative drafting rules.

The rules under examination were those which, having a high technical profile, could be directly applied by regional legislative offices and they were, therefore, considered easy to apply and identify.

The analysis based on the comparison between the application and the failure to apply the rules within a regional law was conducted in successive passages, each one of which corresponding to a qualitative aspect of the law.

The Working Group drew attention to the rules-quality factors which were on the whole applied with greatest recurrence in 39 laws.

---

[17] The group made up of officials from the Tuscany Region (Loredana Balloni, Bruna Berti, Antonella Brazzini, Spartaco Farulli, Domenico Ferraro, Teresa Gottardo, Maria Cristina Mangieri, Massimiliano Mingioni, Antonio Prina and Lucia Silli) is co-ordinated by Dr. Carla Paradiso of the Legislative Quality Service of the Regional Council of Tuscany.

[18] Indice di qualità: la sperimentazione del Consiglio regionale della Toscana. Percorso e metodologia, by the Working Group for the Application of the Unified Drafting Manual, the Application and Monitoring of the Rules Applicable Ex Officio, in the Proceedings of the Seminar on Rules for Drafting Uniformity and Quality Index of Laws: The Experience of the Regional Council of Tuscany. Seminar, Florence, 19 September 2002.

In this way and on the basis of their frequency of application in the sample corpus, a weight was given to the single rules-quality factors (from 1= not important, to 5= very important).

However, in this sector, it appears several requirements cannot be done without [20].

The first is the necessity to use methods and tools which enable reliable and comparable results to be given.

In particular, the following appear to be indispensable:

- tools for the automated recognition of natural language so that the text structures that do not comply with the legislative drafting rules can be identified. These tools are even more necessary for the analysis of extended corpora, on which the work of a large number of specialist is required;
- reliable "metrics" for measuring the errors that are found and the subsequent preparation of these measurement in statistical indexes aimed at expressing the quality levels and, consequently, at making comparisons.

Natural language processing methods, by now consolidated and used also for the recognition of legislative texts, on the one hand, and statistical techniques for checking quality, that are wide-spread in many sectors of production, on the other, can constitute the technical-scientific support for successfully introducing quality control of legislative texts.

The second necessity which seems impelling is to involve and co-ordinate centres of excellence at the highest scientific level in the fields of documentation and legal, linguisitic and statistical processing and the control and evaluation of quality.

In fact, a rigorous and constant discussion amongst specialists in the humanities and sciences is needed, especially if this relationship is consolidated and adds to a long experience, a great opening to the most advanced developments[19].

References

1. S. Baroncelli, S. Faro, Tecnica legislativa e legislazione regionale: l'esperienza delle regioni Toscana, Emilia Romagna e Lombardia, in Iter Legis, 1998, pp. 173-213
2. C. Biagioli, E. Francesconi, P.L. Spinosa, M. Taddei, The NIR Project: Standards and tools for legislative drafting and legal document Web publication, 2003 Proceedings of the International Conference of Artificial Intelligence and Law, Edinburgh, June 24, 2003.
3. C. Biagioli, E. Marinai, P. Mercatali, Documento normativo: note esplicative del DTD per i documenti normativi, in Informatica e diritto, 1, 2000, pp. 55-106;

[19] The place in which these centres of excellence should be seems to be Florence, where, for well-known historical and cultural reasons, the Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG), the prestigious Accademia della Crusca and two University Departments, the Department of Public Law and the Department of Statistics can be found, which are at the height of excellence in their various sectors. After several informal meetings which took place in 2001, these centres have decided, during a meeting held on 23 September 2003 at Ittig, to collaborate together on national and European research projects in this domain.

4. C. Biagioli, Ipotesi di modello descrittivo del testo legislativo per l'accesso in rete a informazioni giuridiche, contribution to the Feasibility Study for the Implementation of the "Access to the Law on the Net", 31 January 2000, pp. 1-90, www.ittig.cnr.it;
5. A. Bolioli, P. Mercatali, F. Romano, Legimatics Methodologies for the Implementation of a Legislative Grammar in atti del Workshop e-Government Modelling Norms and Concepts as Key Issues, Edinburgh 24 June 2003.
6. A. Bolioli, L. Dini, P. Mercatali, F. Romano: For the automated mark-up of Italian legislative texts in XML, in atti del Convegno Jurix 2002, Londra 16-17 dicembre 2002;
7. A. Boer, R. Hoekstra, R. Winkels, T. van Engers, F. Willaert, Proposal for a Dutch Legal XML Standard. (PDF), in Proceedings of the EGOV 2002 Conference (DEXA 2002).
8. M.C. De Lorenzo, Modelli di novelle, in Informatica e diritto, 1, 2002.
9. G. Ferrari, Introduzione alla linguistica computazionale in http://apollo.vc.unipmn.it/~ling_gen/opening.htm;
10. A. Marchetti, F. Megale, E. Seta, F. Vitali, Marcatura XML degli atti normativi italiani. I DTD di Norma in rete, in Informatica e diritto, 1, 2001, pp. 123-148;
11. P. Mercatali., Atti normativi: criteri per l'individuazione e la descrizione, in Technical Report, Internal Memorandum, 2/2003, http://www.ittig.cnr.it/Ricerca/PubRicerca.php;
12. P. Mercatali, Note per la formalizzazione delle disposizioni normative di delega, Technical Report, Internal Memorandum, 4/2003, http://www.ittig.cnr.it/Ricerca/PubRicerca.php.
13. Pietro Mercatali, Legimatica e nessi normativi, Presentation at the National Seminar "Formazione per le tecniche legislative", Turin 17 and 18 June 1999, in Iter Legis, Novembre – dicembre 1999
14. P. Mercatali, Produzione legislativa, informatica e intelligenza artificiale. Sistema sperimentale per il drafting legislativo, in Mercatali P., Soda G., Tiscornia D., Progetti di intelligenza artificiale per la pubblica amministrazione, Franco Angeli, Milan 1996, pp. 84-98.
15. A. Mortara Garavelli, Le parole della giustizia, Einaudi, Turin, 2001;
16. R. Pagano, Le direttive di tecnica legislativa in Europa, Camera dei deputati, Rome, 1997;
17. R. Pagano, Introduzione alla legistica. L'arte di fare leggi, Giuffrè, Milan, 2001;
18. G.U. Rescigno, L'atto normativo, Giappichelli, Torino, 1998;
19. E. Roche, Y. Schabes (eds.), Finite-State Language Processing, The MIT Press, 1997.
20. F. Romano, Drafting legislativo e controllo di qualità: esperienze e proposte, Technical Report, Internal Memorandum, 3/2003, http://www.ittig.cnr.it/Ricerca/PubRicerca.php.
21. P.L. Spinosa, I nomi uniformi dei provvedimenti giuridici adottati da "Norme in Rete", 2003
    III World Congress on Computers and Law, Havana 29 September - 3 October 2003
22. F. Sabatini, Analisi del linguaggio giuridico, in D'Antonio M., Corso di studi superiori legislativi 1988-1989, CEDAM, Padua, 1990;
23. G. Sartor, Riferimenti normativi e dinamica dei nessi normativi, in Il procedimento normativo regionale, Cedam, Padua, 1996.
24. T.M. van Engers, R.A.W. Vanlerberghe, The POWER Light-version: Improving Legal Quality under Time Pressure, in http://lri.jur.uva.nl/~epower (2002).
25. C. Zuliani , Elementi per uno studio dell'evoluzione della tecnica legislativa in Italia dal 1861 al 1986, Parte I, in www.arslegis.it (2003).