Switching from Narrative to Legal Genre

Rodolfo Delmonte and Denise Dibattista
University of Venice

The paper presents a linguistically based approach to text understanding related to texts belonging to legal genre and highlighting some of their properties in the area of semantic and inferential processes. In particular, referential properties of general nouns are discussed and shown at work in an excerpt taken from a EC directive. Proposals to deal with the same problem by Strzalkowski, Cercone, Woods, Brachman are presented, discussed and compared. Our approach is embodied in a system called GETA_RUN presented at various international conferences and freely available from our ftp site.

1. Introduction

This paper deals with the analysis of written texts in legal genre with the aim to highlight linguistic and cognitive differences in its surface and deep form from narrative genre.

Generally speaking, narrative texts deal with entities actually existing and events which actually took place in the real world - disregarding for the moment fairy tales and science fiction stories. In the analysis of narrative texts, the backbone is represented by a temporal sequence or timeline to which the various events making
Switching from Narrative to Legal Genre

up the story are related, and a limited number of topics introduced and then referred
to during the text which attract the focus of attention of reader or listener.

In legal texts, there is neither an internal plot, nor a main protagonist to be
highlighted, nor a temporal progression. Events are all related to a hypothetical
instantiation of some (pre)condition which is the obligatory body of the legal
argumentation, usually also addressed by the title or the heading of the bill or
directive, and there are definitions which express it in more detail.

As to entities, legal texts are characterized by the implicit presence of a universal
quantifier which determines the interpretation of referential expression in terms of
classes of individuals rather than sets or single individuals. For instance, the council
directive we have used to elaborate the computational model concerns the liability of
producers for defective products. However, as we shall see in more detail below,
our ontology is made up of three types of entities: generic entities, classes of
individuals, individuals and sets.

We shall not be dealing with the specificity of the legal lexicon, which however
requires a specialized terminology or in which the common vocabulary is used with a
different meaning from narrative texts.

Switches are placed in the system at all level of computation and we shall discuss
them one by one in relation to each level of representation.

2. Syntactic switches

A switch at syntactic level will typically regard the peculiarities of legal sentential
structures, which might affect the way our parser takes decisions, for instance, when
in presence of conjunctions or punctuation marks; in presence of constituents which
might be regarded both as complements or as adjuncts, like PPs, but when appearing
in a given position in legal texts, they can only be interpreted in a given way;
peculiarities of postmodification, when for instance an implicit participial adjunct is
coordinated with a PP, and so on.

Generally speaking, legal texts are more complex from a structural point of view
than narrative texts. Sentences tend to be very long - in particular definitional ones -
and very heavy stylistically. This is usually due to the need to express complex concepts which are better expressed when connection are kept within the same utterance than when they are expanded at discourse level, intersententially.

Here below is a list of typical problems at a structural level which require a specialized set of rules, activated by a switch:

a. Enumerations or NP lists;
b. PP adjuncts in between auxiliary/modal and main verb;
c. Double subordination;
d. Shall as modality operator;
e. Lack of ‘that’ complementizer deletion that might induce garden paths with NP structure;
f. High number of passive constructions with a systematic omission of the agent that contrary to what happens in narrative texts need not be introduced semantically; also there is a heavy use of passive constructions coupled with modal operators;
g. Premodifiers with restrictive import, rather than as free adjuncts;
h. Ellipsis and use of other implicit subordinate structures like absolute constructions or generic participials;
i. Postmodification with restrictive relative clauses and participials;
j. Parataxis.

At syntactic level, clauses and sentences are connected with a preferred tendency towards parataxis. However, also hypotactic relations are commonly used whenever condensation and precision are the main purposes. Therefore, in a legislative text, the scarce occurrence of connective elements usually coexist with multiple subordinate clauses or sentences introduced in a position unusual in ordinary language.

The use of paratactic structures is functional to legislative language, which tends to explicitness and exhaustiveness whenever possible. Nevertheless, the repetition or juxtaposition of elements may lead to long, sometimes redundant, sentences. Actually, it has been observed that not always the desirability of simplified legislative language is obtained through the use of short sentences.
The problems of semantic connections between sentences, and the necessity of syntactic condensation of legislative texts, are solved, to a great extent, by means of highly nominal constructions, reduced clauses, nominalizations, passives with deleted agents, and embedded sentences.

2.1. Reduced clauses

An important syntactic feature is the prevalence of complex nominal phrases with a high occurrence of postmodification.

Restrictive relative clause is the more frequent case of postnominal modification. This is justified by its function of defining or reducing the meaning of the antecedent noun phrase, a further step towards precision and clarity required by legal genre. For such reasons, restrictive relative clauses are recurrent in legal definitions whenever the legislator intends to limit the class of individuals to whom a rule applies. Furthermore, in legislative texts, these kind of clauses are generally correspondent to conditional sentences:

(1) Any person who imports a product for sale shall be responsible as a producer.

(2) If a person imports a product for sale he shall be responsible as a producer.

It is evident that example (1) is equivalent to (2). The latter is a defining conditional, it is not used to make hypothesis or predictions, but provides the precise meaning of a term which is general and not specific. On the particular form and entailments of conditional clauses we will dwell upon below. Non-restrictive relative clauses are rare in legislative texts and used only when there is no possibility of ambiguity:
(3) The Joint Committee shall be composed on principle of twelve members, five representing the Committee of Ministers and seven representing the Consultative Assembly, the latter number to include the President of the Consultative Assembly, who shall be a member ex-officio.

Example (3) is the only instance of non-restrictive relative clause found in our texts and it is of a simple, not ambiguous interpretation, since the only antecedent of the relative pronoun may be "the President of the Consultative Assembly". As a result, these clauses are largely substituted with coordinate sentences and the repetition of the subject expression is preferred to the relative pronoun.

Postmodification is also realized through non-finite clauses, particularly present and past participles, which emphasize once more, the legislator's attention to condensation. All the instances present the antecedent head identical to the deleted subject of the non-finite verb clause: "the Treaty establishing [...]", "a proposal amending [...]", "damage resulting from [...]"; "regulations issued", "convention ratified", "the rights conferred", etc.

In the examples the -ing construction represents the active voice, whereas the -ed construction is linked with the passive voice. Passives with deleted agents actually are a common means of legal language to give thematic prominence to the element in subject position and therefore achieve depersonalization and abstractness:

(4) Before deciding upon a request mentioned under paragraph 1 above the requested body may, in view of general interests involved, hear the competent public authorities.

The use of premodification with participle is noteworthy: there is hardly any instance of premodification by present participle ("distinguishing feature" is one of the few instances present), while the examples of premodification by past participle ("a finished product", "the injured person", "the proposed measure", etc.) are more common. Since present and past participles maintain the same reference to active and passive voice as in postmodification, this prenominal forms reflect, in their syntactic use, the legislative preference for an impersonal, detached tone.
2.2. Nominalization

A particular case of postmodification is given by prepositional phrases that are used to qualify the head noun. Even though in other genres (narrative, instructional texts, etc.) a noun that specify a preceding phrase is commonly placed in adjective position, legal language opts for postmodifying of-phrases whenever possible:

\[(5)\] The declaration will become effective on the first day of the month following the expiration of a period of three months after the date of its reception by the Secretary General.

A remarkable occurrence of such linguistic behaviour is given by the presence of nominalization, in our case reception in example (5).

Because of its frequency, the tendency towards nominalization of legal language is considered the more relevant lexical and syntactic characteristic. It consists in conveying the functions of the verb to the noun and it is realized in lexicon through deverbal and verbal nouns.

Deverbal nouns are forms obtained by adding to the verbs particular suffixes such as: -\textit{ation} ("presentation", "circulation", "implementation", \textit{etc.}); -\textit{ment} ("arrangement", "apportionment", \textit{etc.}); -\textit{al} ("proposal", "refusal", "dismissal", \textit{etc.}); -\textit{ing} ("stock-farming", "processing", "leasing", \textit{etc.}). In legal genre deverbal nouns have only an abstract meaning, while in other genres (narrative, instructional texts, dialogues) they can be used also for concrete nouns.

It must be observed that this lexical process has clear syntactic properties: the prepositional phrase headed by "of" may correspond functionally either to the subject (e.g. "the functioning of the common market"), or to the object of the transformed verb ("protection of the consumer").

Nominalization has essentially two functions: in the first place it is, together with participial phrases, the principal means of controlling the length of the sentence and obtain condensation. It must be observed that it does not always achieve its purpose since sentences using the nominal style are frequently wordier than the corresponding verbal style.
In the second place, nominalization has a relevant consequence: a noun-like element in a sentence may be moved more freely than a verb and can be placed in subject position. With this lexical device the writer objectifies the content and makes it possible to treat it as something abstracting from specific personal references. This circumstance is in line with the tendency of legislative discourse to do away with the temporal perspective and to achieve generality and abstractness.

2.3. Passive Voice

Passives constructions, may be regarded as a useful means to obtain condensation and depersonalization. However the high frequency of passive occurrence (see ex. 6)) and the complex use of multiple passives, require a further justification.

(6) The provisions of Articles 1, 2, 3 and 4 of this Protocol shall be regarded as additional articles [...]. This Protocol shall be ratified at the same time [...]. The instruments of ratification shall be deposited [...]. Done at Paris.

(7) A declaration made in accordance with this article shall be deemed to have been made in accordance with paragraph 1 of Article 63 of the Convention.

It cannot be denied that passives with deleted agent give to legal genre its typical communicative mode, namely the impersonal and detached tone, but they also contribute to its performative function.

Austin (1962), in his established work on performatives states that a verb in the passive voice is an indubitable performative, which is usually found in formal and legal occasions. Furthermore, the "hereby" criterion, used to detect performativity, is applicable to passives. As Austin suggests, when "hereby" can be inserted, it indicates that the written utterance is the instrument affecting the act of authorising, forbidding, etc. and is therefore performative.
Although passives with deleted agent are found also in conditional clauses, here they cannot have a performative function.

(8) Any member of the Council of Europe may withdraw by formally notifying the Secretary General of its intention. Such withdrawal shall take effect at the end of the financial year in which it is notified. [...] If the notification is given in the last three months of the financial year, it shall take effect at the end of the next financial year.

In accordance with the explanation proposed by Vander Linden (1994) for instructional texts, the use of agentless passives seems to depend on whether the action has been mentioned before or not. If an action or an event has already been introduced in the previous text, it takes the present tense, agentless passive form, as in (8); otherwise a simple present, active form is used.

2.4. Embeddings

Embedding is a device frequently employed in legislative texts to provide condensation and precision. It consists in the insertion of a modifying phrase or clause, often introduced by a subordinating conjunction, within the clause structure. The adverbial function of embeddings is realized by finite-verb clauses, as in example (9); non-finite clauses, as in example (10); and prepositional phrase, as in example (11):

(9) The following may bring a case before the Court, provided that the High Contracting Party concerned, if there is one, or the High Contracting Parties concerned, if there is more than one, are subject to the compulsory jurisdiction of the Court.

(10) The liability of the producer may be reduced or disallowed when, having regard to all the circumstances, the damage is caused both by a defect in the product and by the fault of the injured person.
(11) The same shall apply, in case of an imported product, if this product does not indicate the identity of the importer referred to in paragraph 2, even if the name of the producer is indicated.

Embedded clauses with finite verbs, like in (9), are usually of conditional type and are often placed next to the elements they qualify, thus contributing to the precision of the text.

Nonetheless, other recurrent embeddings are verbless clauses like "if necessary", "if possible", etc., where the assumed subject is an impersonal "it" referring to the main clause as a whole. In this case the text appears heavier and a potential source of ambiguity. In turn, (10) is an embedding with a non-finite clause, which may give rise to problems concerning the identification of its subject, since it does not occur in the main clause.

Lexical-Functional Grammar, the theory that we adopt as our framework, treats this kinds of clause as Closed Adjuncts, which may modify the event described by the main predicate. They are subject to anaphoric or arbitrary control. It means that PRO (called "big pro") the morphologically unexpressed subject of non-finite clauses, has an antecedent in the main clause or, in case no controller is available, it is given arbitrary interpretation (see: Bresnan (1982); Delmonte (1988; 1992)).

In example (10) the antecedent of the PRO is not recoverable from the text and, moreover, the temporal reference is non-specific, we are therefore in presence of the typical conditions that give rise to arbitrary interpretation. In legislative texts this use of non-finite embedded clauses is common, even though it is sometimes considered redundant and confusing for a textual comprehension.

Finally, adverbial modifiers may be used as prepositional phrases as in (11). They have the typical adverbial mobility and so may be inserted quite freely. However, they are usually close to the elements they qualify, thus acting as substitutes for longer clausal structures.

The present analysis of embedded structures suggests a feature of legislative language already mentioned: a constant tension between condensation and precision, which may become a potential source of ambiguity.

Differently from narrative texts, the occurrence of anaphoric links between sentences is scarce: pronouns are found only when no ambiguity may arise and, in
most cases, when their antecedents are in the same sentence; finally, substitution, ellipsis and other reference devices are not common.

Coherence is usually given by the repetition of lexical items, which has the function of making explicit the meaning of a possible reference, therefore avoiding the risk of ambiguity that the use of pronominal anaphors might cause.

3. PP attachment and Parsing

Legal texts are very rich in adjuncts at all levels of syntactic structure, so it is necessary to give a detailed analysis of the mechanisms underlying the treatment of complements and adjuncts from a parsing point of view. Consider now the well-known problem of PP attachment or Syntactic Closure (see Delmonte, 1984), which concerns the way in which Prepositional Phrases modifiers or arguments of a given lexical head should be dealt with in a parsing scheme. In the literature we have two different proposals: one grammatical called "garden-path" (see Frazer) and the other purely semantic called "incremental-interactive" (see Steedman & Altmann); we are only interested in the grammatically based one, which we illustrate briefly here below:

A. The Garden Path proposal has the following main features (see Kennedy et al.,(1989)):

- choose the first available analysis, or words in a sentence are incorporated into complete syntactic structures at the earliest possible opportunity (i.e. the structural description is developed "on-line", word-by-word);
- do not postulate any unnecessary node - attachment of words within a structure is invariably achieved in a way which minimises the number of nodes demanded (the principle of Minimal Attachment - MA);
- if consistent with the rules of the grammar, attach each incoming word into the phrase currently being analyzed - or new words are
incorporated, wherever this is possible grammatically, into the current clause or phrase being processed (the principle of Late Closure - LC).

In a typical minimal pair example such as the following,

(12)  a. She positioned the dress on the rack  
       b. She wanted the dress on the rack

the two principles mentioned above, MA and LC, make conflicting predictions: in (12a) MA predicts that the PP "on the rack" shall not be assigned as adjunct of the head "dress" but as locative oblique argument of the main verb "position", thus complying with the general criteria of economicity and psychological efficiency based on grammatical issues; however the principle LC would make just the opposite prediction, since there is no grammatically motivated criteria to prevent the PP to be computed locally as a semantically compatible adjunct of the head "dress". In version (12b) the two principles will make just the opposite predictions, so that MA would predict wrongly that the PP attaches to the main verb as a locative adjunct or oblique argument, whereas only LC would apply correctly.

It is a fact that examples such as these are genuinely ambiguous cases and there is no way to state general principle which could apply to both equally well, producing the best efficient result. Even if we try to maximixe on the fact that in a. we already know what the Verb Guidance (see Mitchell (1989)) is, i.e. we expect an oblique PP to be present somewhere in the following structure, we still need to take a stance as to whether we rely on a
- selecting procedure, or structure assembly process;
- monitor procedure, or structure checking process;
where the second one may involve rejecting the initial structure, backtracking and reassembling a new structure (see Mitchell, 126).

3.1. The Parser

The parser we use is a rule-based parser, however, being a context-sensitive grammar it incorporates naturally linguistic restrictions which make it particularly
attractive. It is not a principle-based parser in the sense of Chomsky (1986), in that there are no external principles which apply to syntactic representations in order to check their grammaticality. We could dub our parser as a global approach to linguistic parsing, in that linguistic restrictions are integrated in the parsing schema. In this sense, all selectional semantic filtering operations are carried out as soon as possible. This is allowed especially by the overall theoretical framework offered by LFG, which however we enrich, as far as syntactic c-structure and lexical representation are concerned. In particular, LFG already provides a very restricted way to deal with grammatical phenomena because of its modular structure: lexical phenomena are dealt with at a l-structure, where lexical variables are bound; syntactic phenomena are dealt with at c-structure, where syntactic variables are bound; finally, anaphoric phenomena are treated at f-structure level, where pronominal binding takes place. At the same level, also semantic compatibility between adjuncts is computed. We shall discuss the theoretical background in the following Chapter II.

A strict comparison with principle-based parsing as presented in Berwick et al. (1991) is, in our opinion, out of place for two reasons: most of the parsers presented are toy-parsers in the sense that they are not intended for parsing real extended texts; they simply try to show the feasibility of the principle and parameters approach. None of the parser actually implement the theory in a step by step fashion, due to the fact that in order to mimic Move-a in a real parsing schema D-structure should be produced first and then traces should appear where NP movement applies. However, the choice of all the parser builders reviewed in the book is to work directly on some version of S-structure. A secondary but nonetheless important feature of principle-based parsers is that, with the notable exception of Dorr's parser, all other parsers work on a two-stage mode, i.e. the structure building process is kept separate and precedes the interaction with universal principles. Dorr's parser is a modular system which tries to make use of semantic restrictions, such as thematic roles constraints on arguments of a given predicate. However, as the author herself admits, the system is still too much "syntactically driven", thus failing to capture a number of structural ambiguities of semantic nature (see Dorr, 179). X-bar theory in itself and the configurational approach to sentence and utterance analysis is deficient in its ability to solve structural ambiguities due to attachment of adjuncts. All major projections, NP, PP, AP, QP, CP, can either constitute an (optional) argument or an adjunct: in turn, adjuncts, being syntactically independent and unrestricted by either
subcategorization or argument structure, require the presence of a semantically consistent network of compatibility tests to be applied before attachment can be decided. As a matter of fact, any grammatical theory in itself, be it LFG or GB/Barriers or GPSG, lacks adequate principles for interpreting the role played by adjuncts in the structure building process, simply because adjuncts are only semantically and not syntactically restricted. However, it is clear to us, that the solution to the problem of adjuncts within systems like LFG or GPSG is much more perspicuous, simply because they are feature-based systems. Some of the solutions adopted in our system are discussed in this chapter.

Here we would like to comment on the most attractive computational features of our parser. Being a DCG, the parser is strictly a top-down, depth-first, one-stage parser with backtracking: differently from most principle-based parsers presented in Berwick (1991), which are two-stage parsers, our parser computes its representations in one pass. This makes it psychologically more realistic. The final output of the parsing process is f-structure which serves as input to the binding module and logical form: in other words, it constitutes the input to the semantic component to compute logical relations. In turn the binding module may add information as to pronominal elements present in the structure by assigning a controller/binder in case it is available. As to the most important features of DCGs, we shall simply quote from Pereira and Warren (1980) conclusions, in a comparison with ATNs:

"Considered as practical tools for implementing language analysers, DCGs are in a real sense more powerful than ATNs, since, in a DCG, the structure returned from the analysis of a phrase may depend on items which have not yet been encountered in the course of parsing a sentence. ... Also on the practical side, the greater clarity and modularity of DCGs is a vital aid in the actual development of systems of the size and complexity necessary for real natural language analysis. Because the DCG consists of small independent rules with a declarative reading, it is much easier to extend the system with new linguistic constructions, or to modify the kind of structures which are built. ... Finally, on the philosophical side, DCGs are significant because they potentially provide a common formalism for theoretical work and for writing efficient natural language systems." (ibid, 278).

Grammatical principles are located within the parser at various levels, in force of the principle that as soon a principle becomes applicable it is instantiated. LFG makes
use of X-bar notation but not in a principled way: for instance, there is no Extended Projection Principle which requires the presence of a Subject NP in the IP level representation. In languages like Italian, IPs may be rewritten simply by VP in case the SUBJect NP is computed in VP internal position, as happens with unaccusatives. Beside, for a maximal projection to be instantiated at c-structure level, the lexical head must be present. Spec-Head agreement is performed very smoothly by means of Prolog internal mechanism of variable instantiation.

Both categorial and subcategorization information is used to guide the parsing process deterministically: in particular, subcategorization both for verbs and adjectives is accessed from the lexical form of the predicate as soon as it is available. When the verb is analysed, syntactic categories like transitive, unaccusative, psychic etc. are used to guide the construction of the VP. As for adjectives, their Pred is inspected in order to decide whether to look for a given complement, which can be a constituent like a PP or a VP in the input string. Subsequently, semantic features may be used to further select these constituents as being appropriate complements to the head.

Selectional restrictions in the form of inherent semantic features associated to any lexical category are used throughout the parsing process. In the building of an NP, a PP is accepted as possible argument or adjunct in case semantic compatibility tests are passed. These tests are performed before the complete structure of the NP is built, and other constituents may be present in the final structure. When building a relative adjunct, semantic features of the head noun are percolated into the embedded open proposition in order to speed up the search for the controlled element, or landing site. This is paramount in languages like Italian, where SUBJect NPs may be freely inverted in postverbal position.

Generally speaking, the same semantic testing procedures are used for adjuncts: they are tested locally for compatibility. This is due to the fact that adjuncts in Italian may be freely interposed between any major constituent and the verb, as well as between the verb and the OBJect NP, differently from what happens in English where strict adjacency is respected (see Delmonte (1987)). We shall concentrate only on the PP attachment problem and basically on the way in which we cope with the need to pass information which has been acquired in a previous step of computation and use it to guide the parsing process.
3.2. *Two mechanisms at work*

Suppose we have to use the information that "position" is a verb which requires an oblique PP be present lexically in the structure, as results from a check in its lexical form. We take the verb in I position and then open the VP complement structure, which at first builds a NP in coincidence with "the dress". However, while still in the NP structure rules, after the head has been taken, a PP is an option freely available as adjunct.

In our parser we always give priority to what is locally possible and thus we favour a "selecting or structure assembly procedure", rather than the second procedure which would allow for backtracking, seen that this is a very risky and time-consuming option. What kind of mechanism do we incorporate in order to make the right decision also in the second example, where the main verb "want" does not guide any argument requirement in the following structure building process.

We have implemented two look-ahead mechanism which are used in the PP building rule and are always triggered, be it from a position where we have a noun as head and we already built part of the corresponding constituent structure; be it from a position where we have a verb as head and we want to decide whether our PP will be adequate as argument rather than as adjunct - in the latter case it will become part of the Adjunct Set.

**Mechanism 1**
- *Cross Compatibility Check*

This mechanism requires the head semantic features or inherent features to be checked against the preposition, which in turn activates a number of possible semantic roles for which it constitutes an adequate semantic marker. For instance, the preposition "on" is an adequate semantic marker for "locative" semantic role, this will cause the compatibility check to require the presence in the governing heading of inherent or semantic features that allow for location. A predicate like "dress" is computed as an object which can be assigned a spatial location, on the contrary a predicate like "want" is computed as a subjective intensional predicate which does not require a spatial location. However, in order to take the right decision, the CCC must be equipped with the second mechanism we implemented;
Mechanism 2

- Argument Precedence

The second mechanism we implemented is a global or metavariable which allows the parser to satisfy the subcategorization requirements in any constituent it finds itself at a given moment if the parsing process - relatively only to PPs though. Suppose that after taking "positioned" as the main verb, this mechanism is activated, by simply copying the requirements on PP oblique locative present in the lexical form associated with the predicate "position" in the lexicon, in the AP metavariable. As soon as the NP "the dress" is opened, after taking "dress" as N at the head position, the parser will meet the word "on", which allows for a PP adjunct. While in the P head position, the parser will fire the CCC mechanism first to see whether the preposition is semantically compatible, and in case it is, the second AP mechanism will be fired. This will cause the system to do the following steps:

i. check whether the metavariable is empty or not;
ii. and in case it is instantiated, to control the semantic role associated with it;
iii. to verify whether the P head is a possible semantic marker for that semantic role: in our case, "on" is a possible semantic marker for "locative" semantic role;
iv. finally to cause the parser to fail on P as head of a PP adjunct of the head noun;
v. produce a closure of NP which obeys Minimal Attachment principle.

3.3. Some examples

In our text there is a great number of examples which can be used as empirical evidence for the need to use lexical information in order to reduce parsing loads resulting from structure monitoring, or rather backtracking procedures. Our examples are taken from the text included as an Appendix at the end of the paper: we mark decision points with a bar:
Council directive of July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products.

At the first boundary we have "of" which is non semantically marked and no prediction is available, so that the default decision is to apply Late Closure, which turns out to be the correct one. When the second preposition is found we are in the NP of the PP headed by "of", and we have taken the date "1985": this will cause the CCC to prevent the acceptance of the preposition "on" as a semantically compatible marker thus preventing the construction of the NP headed by "approximation".

Notice, that in case that would be allowed, the NP would encompass all the following PPs thus building a very heavy NP: "the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products". In case the parser had a structure monitoring strategy all this work would have to be undone and backtracking would have to be performed. Remember that the system does not possibly know where and how to end backtracking unless by trying all possible available combination along the path. In our case, the presence of a coordinate structure would render the overall process of structure recoverability absolutely untenable.

Another important decision has be taken at the boundary constituted by the participial head "concerning": in this case the CCC will take the inherent features of the head "States" and check them with the selection restrictions associated in the lexical form for the verb "concern". Failure in this match will cause the NP "the Member States" to be closed and will allow the adjunct to be attached higher up with the coordinated head "laws, regulations and administrative provisions". In this case, all the inherent features are collected in a set that subsumes them all and can be used to fire CCC.

Notice that the preposition "for" is lexically restricted in our representation for the noun "liability", and the corresponding PP that "for" heads interpreted as a complement rather than as an adjunct. We include here below the relevant portion of each utterance in which the two mechanisms we proposed can be usefully seen at work. We marked with a slash the place in the input text in which, usually when the current constituent is a NP a decision must be taken as to whether causing the parser
to close (MA) or to accept more text (LC) is actually dependent upon the presence of some local trigger. This trigger is mostly a preposition; however, there are cases in which, see e., f., h., i., the trigger is a conjunction or a participle introducing a reduced relative clause. Coordinate NPs are a big source of indecision and are very hard to be detected if based solely on syntactic, lexical and semantic information. For instance, e. can be thus disambiguated, but h. requires a matching of prepositions; In the case represented by i. we put a boundary just before a comma: in case the following NP "the Member State" is computed as a coordination - which is both semantically, syntactically and lexically possible, the following sentence will be deprived of its lexical SUBject NP - in this case, the grammar activates a monitoring procedure independently so that backtracking will ensue, the coordinate NP destroyed and the comma computed as part of the embedded parenthetical (which is in turn an hypothetical within a subordinate clause!!). Notice also that a decision must be taken in relation to the absolutes headed by a past participle which can be intended as an active or a passive past participle: in the second case the head noun would have to be computed as an OBJect and not as a SUBject

b. a differing degree of protection of the consumer | against damage caused by a defective product | to his health or property

c. in all member states | by adequate special rules, it has been possible to exclude damage of this type | from the scope of this directive

d. to claim full compensation for the damage | from any one of them

e. the manufacturer of a finished product, the producer of any raw material or the manufacturer of a component part | and any person

f. The liability of the producer | arising from this directive

g. any person who imports into the community a product | for sale, hire or any form of distribution | in the course of his business

h. both by a defect in the product | and by the fault of the injured person
However, if... the commission does not advise the Member State concerned that it intends submitting such a proposal to the council, the Member State

4. Semantics and Discourse Model

The main tenet of the theory supporting the construction of our system is that it is possible to reduce access to domain world knowledge by means of contextual reasoning, i.e. reasoning triggered independently by contextual or linguistic features of the text. It is sensible to assume that when understanding a text a human reader or listener does make use of his encyclopaedia parsimoniously. Contextual reasoning is the only way in which a system for Natural Language Understanding should tap external knowledge of the domain. In other words, a system should be allowed to perform an inference on the basis of domain world knowledge when needed and only then. In this way, the system could simulate the actual human behaviour in that the access to extralinguistic knowledge is triggered by contextual factors independently present in the text and detected by the system itself.

It is also our view that humans understand texts only whenever all the relevant information is supplied and available. Descriptive and narrative texts are usually self-explanatory - not so, literary texts - in order to allow even naive readers to grasp their meaning. Note that we are not here dealing with spoken dialogues, where a lot of what is meant can be left unsaid or must be implicitly understood.

In the best current systems for natural language understanding, linguistic components are kept separate from knowledge representation, and work which could otherwise be done directly by linguistic analysis is duplicated by the inferential mechanism. Linguistic representation is usually mapped onto a logical representation which is in turn fed onto the knowledge representation of the domain in order to understand and validate a given utterance or query.

Thus the domain world model must be priorly built, usually in view of a given task the system is set out to perform. However, it is clear that this modelling is
domain and task limited and no generality whatsoever is achieved from it. In some of these systems, the main issue is how to make the two realms interact as soon as possible in order to take advantage of the inferential mechanism to reduce ambiguities present in the text or to allow for reasoning on linguistic data, which otherwise couldn't be understandable.

We assume that an integration between linguistic information and knowledge of the world can be carried out at all levels of linguistic description and that contextual reasoning can be thus performed on the fly rather than sequentially. This does not imply that external knowledge of the world is useless and should not be provided at all: it simply means that access to this knowledge must be filtered out by the analysis of the linguistic content of surface linguistic forms and the abstract representations of the utterances making up the text. One of the major problems in every attempt at real text understanding has been that of controlling the inferences associated with the interpretation process. In our opinion as well as in Partee's and Dahlgren's opinion, separate semantic levels - lexical level, grammar and parser, pronominal and discourse level, temporal level - conspire to produce the global interpretation. Indeed, discourse analysis requires a highly flexible structure where each individual source of semantic knowledge contributes its own local inference which is then combined at a higher level. In this perspective, semantic interpretation is carried out incrementally and the whole process requires cooperation among separate semantic modules in order to produce a coherent inferential process. Such an approach makes the reasoning process both tractable and computationally efficient.

In particular, as in Dahlgren's approach, we assume that an autonomous syntactic component draws upon lexical semantic information for tackling problems such as prepositional phrases attachment, constituency interpretation and its coupling to argument functions or adjunct functions, and word sense disambiguation. The remaining components, however, in our system do not work in parallel, but are fed by the output of the parser, onto which pronominal binding and quantifier scoping is carried out, also using semantic interpretation. The remaining upper modules are essentially semantic interpretation modules including temporal reasoning where spatiotemporal inferential processes are fired; discourse topic hierarchy construction to enable reference resolution at discourse level (see Delmonte & Bianchi, 1994)), and finally the building of a discourse model of facts.
Semantic peculiarities concern genre specialization that affect the way in which the interpretation process is tied up with linguistic forms. It is a fact that linguistic structure and form is the first means by which meaning is conveyed to language users, and it will be shown that it is often contingent upon genre.

Generally speaking, building a Discourse Model is a precondition for any reader or generic addressee of the contents of a legal text to enable reference to entities and events contained in the text. A DM is clearly only a part of the overall process of understanding which makes heavy use of background mutual knowledge on the side of the addressee in order to carry out the complex inferences required by this genre.

In line with current assumptions about the ontological status of entities and events referred to in any legal texts, we also assume that they are spatiotemporally disanchored and they have no counterpart in the real present world: in other word, their referential status is abstracted from spatiotemporal restrictions. Nonetheless, the DM will be represented as a list of facts and sits very much in the same way in which it would happen in narrative texts. The reason for this is very simple, the DM constitutes the semantic informational representation of the linguistic structure of any text or discourse. It aims at simulating the processes underlying anaphora and reference resolution within the text, thus registering and storing information in a given formal format which is the actual interpretation related to entities and events contained in the text.

For instance, a norm, directive or other legal text might contain Obligations which are expressed by a specialized use of the modal “shall” and as is the case with Permissions which are highlighted by the presence of the modal “may”, are treated as facts in the DM; in turn, Hypotheses do not count as extensional objects and are carefully set apart from Conditions which assume a similar syntactic pattern but have a different semantic marker.

If we take reference to classes as the norm in legal language we still must allow for individuals - the current directive is one such case - or sets, the commission or the European community, or still the Member states.
5. Reference and General Nouns

As for human entities, we want to differentiate generic from defining descriptions or properties; in turn these descriptions, which might be exemplified by "consumer" or "producer" in our case, are roles attached to default properties or general nouns like "person" or "man" which are computed as generic entities. A side effect of the special use of referential expressions in legal texts is their lack of referential content: in other words, legal texts do not refer extensionally to entities in the world but deal with generic entities or with classes of individuals; proper nouns may only refer to institutions and as such they can be treated as collective entities or sets; common nouns may deal with abstract concepts belonging to the vocabulary of law or to concrete individuals or objects which could be found in the world but needn't do so. Usually, when common nouns are introduced in a directive or bill they are individuated by a specific definition: this is particularly so, when those nouns are the addressee or the main topic to which the norm applies.

The general noun "person" should be referentially empty in legal contexts: consider its lack of gender specification and its inherent generic nature if compared to common nouns related to social roles. It is heavily used to refer either to actual or to potential addressees: in particular, we might be talking about "producers" and "consumer", and then go on talking about "the injured person" or "two or more persons" which can cause damage. In fact, "person" can be used both with modifiers and without them, as a definite NP or as an indefinite one. When it is used as a definite NP, it requires some modifier to be properly corefered; no bare NPs can be created in association with such a generic head.

While common nouns may refer either to individuals or to classes, singular definite NPs with a generic head cannot be computed as individuals but only as classes and this must be inferred from the modifier or adjunct that specifies it adequately.

Plural definite NPs with generic heads are computed as classes, and can be used to corefer to other singular generic nouns or to specific nouns which they share default properties with.
Differently from what happens in narrative texts, legal texts introduce indefinite NPs not to assert the existence of some individual or set but to refer to classes or to generic entities.

Strzalkowski and Cercone (1989) in their work introduce the problem of reference from a semantic computationally feasible point of view. This is how they define their proposal:

"... we introduce a layered model of reality (the universe) as perceived by a discourse participant, and define relative singularity of objects in this universe as an abstraction class of the layer-membership relation. Subsequently, linguistic descriptions and names are classified as singular, measurably singular, or non-singular depending upon what they are assumed to denote in the universe. The relationship between objects referred to in discourse and classified into different layers (levels) of the universe model has a particular significance for resolution of certain types of cohesive links in text. We call these links remote reference because they cross level boundaries." (p.172)

The tripartite classification of entities in the model reflects our proposal where we have generic entities, classes and individuals. As in S & C they are partially ordered by the relation "more informative than": intuitively, they introduce a relation of relative singularity among objects that allows us to break down the universe of objects into classes or levels, where a lower level L0 consist of manifestations, extensions, instances of objects at level L1. In fact S & C always speak of the existence of three main levels, one for each type of denotation: if L0 contains objects with a measurably singular interpretation, L-1 may contain the same object with a specific or singular interpretation - if possible at all, and L+1 will contain referring expressions - nouns, definite descriptions, pronouns but also other nominals denoting objects which are non-singular. If level L+1 contains generalizations of objects from L0, then level L-1 will contain their specializations of extensions (ibid., 177). Any non singular object can potentially be decomposed into instances in many ways, depending upon the relation that bind the two objects, or coordinate - part_of, instance_of, genus, specimen etc.; also, and more generally, whenever a higher-level
object is decomposed with two different coordinates, the resulting sets of instances need not belong to the same level (ibid., 178).

However, we will assume that natural language users, when introducing referring expressions in a text or discourse will abide the wellknown maxim that requires them not to be ambiguous. In particular, in our model the spatiotemporal location indeces assigned according to semantic and linguistic criteria are the only possible coordinates but cannot be used to partition our discourse model into layers or levels, with respect to some current level, corresponding to the level of reference at a present point of discourse. We do not find any convenience in introducing such layers into the overall computation of entities and their reference. In turn S & C do not describe in detail how one could compute remote reference in discourse, apart from establishing an obvious definition that relates two objects by means of remote reference in case they have different kinds of singularity. Naming an object previously existing in the discourse model may end up simply with a coreferring operation, or with the need to create a supercontext in case it is less informative than the existing entity, or a subcontext in case it is an instance of the previously existing entity (ibid., 182).

Now consider the problem of general nouns: person, man, individual etc. They constitute the most common and most frequent noun in our texts and this is due to their referential properties: they can pick the maximum set of human entities available when used with no attribute or modifier; they can also be used to corefer to single classes of individuals by adequate modification. In other words, the concept “person” will be present at different levels of generality in the DM and it will be instantiated with a different semantic identifier according to its “dattrs” (see Woods and Brachman). Reference to one or the other of the instantiations of “person” will depend upon the way in which the discourse entity is described, in other terms on its “dattrs”, i.e. attributes, properties, parts, constituents, features, relations and so on. Notice that in our DM every entity has a description in terms of situation semantics, with a polarity and a spatiotemporal anchorage in terms of indices for main spatial and main temporal location.
6. Inferential processes

In our system, the problem of assigning properties to entities in narrative genre is carried out in two main phases: first only referring expressions (ref_exs) asserted as best candidates for topichood and ranked as Main, Secondary, Expected or Potential Topics by the Topic Hierarchy Mechanism are considered. This might or might not apply to legal genre, on the basis of the assumption that there is no plot nor main character to focus on. However, legal texts contain pronominials which refer in the previous discourse segment and cannot be done away with. In case no such coreferent is found, the remaining DM is searched for. Generally speaking, there are four possible ways in which the problem might be solved:

CASE A
there is already at least one such entity in the Discourse Model (DM) with the same class restriction which could be used to corefer or cospecify the current ref_ex disregarding its possible contextually determined properties and considering only the class or name that has been used as first or Initial Description (ID);

CASE B
the entity to be picked up as possible coreferent or cospecifier must be semantically equal: i.e. we consider number of the current ref_ex and if plural, check whether the entity has the same cardinality or is asserted as a class; and if singular look only for individuals;

CASE C
the entity and the ref_ex must have the same attributes, roles or other properties assigned to it in the DM and present in the current ref_ex;

CASE D
works as case C and in addition it looks only for properties associated to entities asserted as Topics of discourse in the previous text; in case no such property is present, a new entity is entered in the DM.
It is easy to notice that, case D is the most restrictive but also the most genre bound procedure: in fact, it is only useful in case the text is a narrative one.

The first question is discovering whether the current ref_ex is already present in the discourse model. If the problem at hand were that of matching the current linguistic form of the predicate or concept with those present in the discourse model (DM) this would be easily solved. However this procedure is clearly insufficient and leads to mistaken matches in case cardinality, definiteness, or simply the modifiers of the current linguistic use of a certain entity present in the DM does not coincide semantically with it.

The overall framework is further complicated by the presence of generic entities and by the fact that what we are dealing with are usually entities which might be easily subsumed by them or constitute a specific subclass. A “consumer” or a “producer” is clearly such a subclass in case the general noun “person” has already been introduced in the DM. However, problems will arise when the contrary applies, i.e. whenever a general noun is introduced with specific properties added by adjunction or modification: in this case it would both represent a subclass of a generic entity and a coreferent of a some class of individuals or role.

The general problem to be solved might be coped with by the following procedure:

1. find all semantic identifiers associated in the DM with the predicate of the current referring expression starting from the most recent ones in the stack of ids’ associated with some previous topic;
   1.i search all ids’ starting from the most recent one and make a list;
   1.ii then for each id, look whether the current predicate head is associated to some property in the DM;
   a.1 no identifier associated with the head, goto 2.
   a.2 there is at least one identifier associated with the head;
      b. verify whether the set of adjuncts, modifiers and other property functional assigners, MODS associated to the current head is empty or not
      b.1 if it is empty, goto f.;
      b.2 if it is nonempty, find all predicate heads associated to MODS;
c. search the DM for properties associated to the predicate head other than the head itself and the default properties;
c.1 if no property is found in the DM goto f.;
c.2 if the list of properties in the DM is nonempty:
d. make the intersection between the two sets of properties: set1 from MODS, set2 from DM;
d.1 if the intersection is empty goto e.;
d.2 if the intersection is nonempty and the property/ies found is/are equal to the one/s contained in set1 then fail;
d.3 else continue.
e. check in the external knowledge of the world whether the two sets contain properties which are synonymous or which are inferentially derivable from other properties;
e.1 if there are some such properties, remove them from the list;
e.2 if not continue;
f. check preliminarily whether the two sets of properties are nonempty; then, check number, and if plural check cardinality and then
f.1 if it is the same assert the set of properties in MODS;
i. a singular is compatible with inds;
ii. a plural is compatible with sets and classes
• in legal genre, a singular is compatible both with singular and with classes!!
f.2 if it is different fail;
g. search for other ids associated to the current head;
g.1 if no other id is present goto end;
g.2 if some other id is present goto b.

2. find all semantic identifiers associated in the discourse model (DM) with the predicate of the current referring expression which are not included in the list of the topic identifiers;
2.i if it is empty, goto 3.
2.ii if it is nonempty goto b.
3. find all semantic identifies associated in the External Knowledge Base with the predicate of the current referring expression;
   3.i if it is empty, goto end.
   3.ii if it is nonempty goto b.
end.

Let's now go back to general nouns: in any text, "person" should be introduced effectively only whenever a number of possible specific entities which might be subsumed by "person" already exist in the DM. In the texts we analyzed, this is what happens: in particular, the Council Directive we used and implemented which is discussed in mode detail at the end of this paper, has the following textual structure:

1. introduce main topics and addressee of its contents, which are "producers" and "consumers" living in the "Member States"; the directive concerns the protection of consumes and the producers' liability in relation to injury or damage which might result from defective products. Importers of products who present themselves as producers are also regarded as such.
2. subsequently, both main topics are coreferred by means of the general noun "person".

On first appearance of common nouns which might be subsumed by "person", the system checks whether there is any such entity in the DM: in the affirmative case it simply inherits its identifier and in the negative case, a generic entity is asserted in the DM. This entity is not an individual nor a set but has the following properties:
- it may subsume other generic entities of the same kind;
- it may be used to infer the nature of references to individuals by means of "person" and some specific attribute or property.

In particular, in the following text we find "several persons liable" and "injured person". The first referring expression is computed as "producer" and the second one as "consumer". In order to get this interpretation, which is the one intended in the Directive, an inferential process must be carried out on the basis of external knowledge of the world. However, the trigger to start this process is constituted by the attribute "injured" which is computed in the property checker algorithm as a
possible new property to be associated to some entity already existing in the DM. The search starts from a referring expression that has some attribute or modifier which is a property not yet asserted for the corresponding entity. Consider the case of "injured person", the algorithm by looking into the DM recovers the identifier associated to "consumer" on the basis of the inference that he is a "person" which is expressed in the following list of facts:

\[
\begin{align*}
\text{class}(\_\text{, Id}) \\
\text{class}(\_\text{, Ind}) \\
\text{fact}(\_\text{, isa, }[\_\text{:Ind, class:person}], 1, \_\text{, }\_\text{)} \\
\text{fact}(\_\text{, ist_of, }[\_\text{:Ind, class:man}], 1, \_\text{, }\_\text{)} \\
\text{fact}(\_\text{, role, }[\text{consumer, Id, Ind}], 1, \_\text{, }\_\text{)} \\
\text{fact}(\_\text{, isa, }[\_\text{:Id, class:consumer}], 1, \_\text{, }\_\text{)} \\
\text{in}(\_\text{, Id, Ind})
\end{align*}
\]

When the Id of “consumer” is recovered from the DM by the property checker that looks for correspondences between literal predicates and their relations in the DM, the “injured” attribute and the Id is passed on to the following inference engine that collects knowledge of the world associated to the trigger and checks to see whether it applies to the current identifier and its properties in the DM.

\[
\text{infer_process(Trigger, Id):-} \\
\text{infer_trig(Trigger, Props),} \\
\text{infer_rels(Id, Props).}
\]

This is done by means of an inferential process that takes as input external knowledge of the world, where we deposit information related to implicit knowledge, mutual knowledge and specialized information which could be part of T-box component in a knowledge base understanding system. In those inheritance networks (see Woods (1978)), each concept that the system understands is represented in a network of concepts, which have to cope with the problem of internal recursion - i.e. of a concept defined in terms of another concept. Our taxonomy is only a small list of facts and the portion that interests us now is represented as follows:
infer_trig(injured, [cause(damage, 0), has(protection, 1), liable(Id, 0)]).
infer_trig(liable, [cause(damage, 1), has(liability, 1)]).

Each term is made up of a trigger, "injured", "liable" which are the properties by means of which the general noun "person" is modified; in turn each trigger is associated with a set of relations and properties which have as argument either a class predicate and a polarity, or a variable and a polarity.

In the text, we learn that "injured person" is only used to corefer to "consumer"; we also know that in order to be interpreted as "injured" a person has not to have caused the damage, nor to be "liable" for it, and finally be the one that is given "protection". By definition, then, a person is liable in case "he caused damage" and in case "he has liability".

infer_rels(Id, [Prop(Id, 1)|Props]):-
    fact(_,Prop, [_:Id], 1, _, _),
    infer_rels(Id, Props)
  ;
infer_rels(Id, [Prop(Id, 0)|Props]):-
    (not fact(_, liable, [_:Id], 1, _, _))
  ;
    fact(_, liable, [_:Id], 0, _, _))
)  
infer_rels(Id, Props)
  ;
infer_rels(Id, [Rel(Prop, 1)|Props]):-
    fact(_, Rel, [_:Idx, _:Id], 1, _, _),
    fact(_, isa, [_:Id, class:Prop], 1, _, _)
    infer_rels(Id, Props)
  ;
infer_rels(Id, [Rel(Prop, 0)|Props]):-
    fact(_, Rel, [_:Idx, _:Id], 0, _, _),
    fact(_, isa, [_:Id, class:Prop], 1, _, _)
).
7. DM and inferences: an example

We will show here below the DM relatively only to three entities: consumer, producer and their general noun "person" subsuming both. In utterance 4 with first appearance of "consumer and producer" the system generates a generic entity "person" which subsumes both: the inclusion relation is registered by the fact in (_, A, B) where A is the semantic identifier or initial description of the superset or superobject, and B is the subsumed entity:

[Whereas approximation of the laws of the 'Member' 'States' concerning the liability of the producer for damage caused by the defectiveness of his products is necessary, because the existing divergences may entail a differing degree of protection of the consumer, against damage caused by a defective product to his health or property.]

ent(infon102, id22)
fact(infon103, isa, [arg:id22, class:person], 1, univ, univ)
class(infon104, id23)
class(infon105, id24)
fact(infon106, cause, [causer:id5, arg:id24], 1, univ, univ)
fact(infon107, inst_of, [ind:id24, class:abstract_state], 1, univ, univ)
fact(infon108, isa, [ind:id24, class:damage], 1, univ, univ)
fact(infon109, against, [arg:id23, malefactive:id24], 1, univ, univ)
fact(infon110, inst_of, [ind:id23, class:man], 1, univ, univ)
fact(infon111, isa, [ind:id23, class:consumer], 1, univ, univ)
in(infon112, id23, id22)
fact(infon113, role, [consumer, id23, id22], 1, univ, univ)
fact(infon114, cause, [arg:id4, damage:id24], 1, univ, univ)
class(infon115, id25)
fact(infon116, inst_of, [ind:id25, class:man], 1, univ, univ)
fact(infon117, isa, [ind:id25, class:producer], 1, univ, univ)
in(infon118, id25, id22)
fact(infon119, role, [producer, id25, id22], 1, univ, univ)
class(infon133, id29)
fact(infon134, differing, [ind:id29], 1, univ, univ)
class(infon135, id30)
fact(infon136, has, [experincer:id23, arg:id30], 1, univ, univ)
fact(infon137, inst_of, [ind:id30, class:abstract_state], 1, univ, univ)
fact(infon138, isa, [ind:id30, class:protection], 1, univ, univ)
fact(infon139, of, [arg:id29, specif:id30], 1, univ, univ)
fact(infon140, inst_of, [ind:id29, class:[measure]], 1, univ, univ)
fact(infon141, isa, [ind:id29, class:degree], 1, univ, univ)

In utterance 6 we learn that the producer should be made liable at certain conditions, and this is registered as an additional fact about producers:

[Whereas protection of the consumer requires that all the producers involved in the production process should be made liable in so far as their finished product, component part or any raw material supplied by them was defective.]
fact(infon248, liable, [nil:id25], 1, univ, univ)

In utterance 8 the text introduces an undefined set of “several persons” which are liable for a certain damage and are related to protection of the consumer in the same context. The system assigns to “injured person” the same identifier as the consumer, creates a new class of entities “person” which are liable and cause “damage”, included in the same superset of producers:

[Whereas, in situations where several persons are liable for the same damage the protection of the consumer requires that the injured person should be able to claim full compensation from any one of them.]
fact(infon323, injured, [ind:id23], 1, univ, univ)
class(infon324, id73)
fact(infon325, full, [ind:id73], 1, univ, univ)
fact(infon326, inst_of, [ind:id73, class:legal], 1, univ, univ)
fact(infon327, isa, [ind:id73, class:compensation], 1, univ, univ)
fact(infon328, cause, [arg:id25, damage:id24], 1, univ, univ)
class(infon329, id74)
fact(infon330, liable, [nil:id74], 1, univ, univ)
fact(infon331, cause, [arg:id74, damage:id24], 1, univ, univ)
in(infon332, id25, id74)
fact(infon333, inst_of, [ind:id74, class:man], 1, univ, univ)
fact(infon334, isa, [ind:id74, class:person], 1, univ, univ)
class(infon335, id75)
fact(infon336, inst_of, [ind:id75, class:abstract_state], 1, univ, univ)
fact(infon337, isa, [ind:id75, class:situation], 1, univ, univ)
fact(id76, claim, [experiencer:id74, theme_aff:id73, source:id74], 1, tes(fin01_aq4), univ)

In utterance 10 liability for damage is charged on the producer: the system looks for a similar predication in the knowledge base and finds the one related to infon248, which was asserted in utterance 6 above:

[The producer shall be liable for damage caused by a defect in his product.]
fact(infon382, cause, [arg:id25, damage:id24], 1, univ, univ)
class(infon383, id84)
fact(infon384, in, [arg:id84, locative:id5], 1, univ, univ)
fact(infon385, inst_of, [ind:id84, class:legal], 1, univ, univ)
fact(infon386, isa, [ind:id84, class:defect], 1, univ, univ)
fact(infon387, cause, [causer:id84, arg:id24], 1, univ, univ)
fact(id85, be, [prop:infon248], 1, tes(f3_dd07), univ)

In utterance 11 we find “the injured person” again, and the system picks up id23 associated to consumer. Notice the computation of the meaning for “relationship” which has “causal” as modifier: the system understands it as a relation and finds infon387 which is associated to a fact in the knowledge base asserted in the previous utterance and saying that there is a “cause” relation between “defect” and “damage”. This was expressed in terms of semantic roles, i.e. the defect is interpreted as a “causer” and its argument is id24, the “damage”. In the new utterance, this is linguistically formulated in terms of “relationship”, where there is a semantic marker “between” which expresses a relation, and the relation has two arguments:
[The injured person shall be required to prove the damage, the defect and the causal relationship between defect and damage.]

ind(infon423, id95)
class(infon424, id96)

fact(infon425, inst_of, [ind:id96, class:abstract_state], 1, univ, univ)
fact(infon426, isa, [ind:id96, class:defect_damage], 1, univ, univ)
fact(infon427, between, [arg:id95, relation:id96], 1, univ, univ)
fact(infon428, cause, [arg:id95, relation:infon387], 1, univ, univ)
fact(infon429, inst_of, [ind:id95, class:abstract_state], 1, univ, univ)
fact(infon430, isa, [ind:id95, class:relationship], 1, univ, univ)
fact(infon433, [defect, damage, relationship], [arg:id97], 1, univ, univ)
fact(id98, prove, [actor:id23, prop:infon433], 1, tes(finf1_dd07), univ)
fact(id100, require, [actor:id23, prop:id98], 1, tes(f4_dd07), univ)

In utterance 12 the same predication is present, "be liable" and the same infon248 is picked up:

[Where, as a result of the provisions of this directive, two or more persons are liable for the same damage.]

fact(id104, be, [prop:infon248], 1, tes(f49_dd08), univ)

In utterance 13 a pronoun is introduced intersentially to corefer to the same persons, associated to the class of producers:

[They shall be liable jointly and severally, without prejudice to the provisions of national law concerning the rights of contribution or recourse.]

fact(id109, be, [prop:infon248], 1, tes(f3_dd10), univ)

In utterance 14, we find a reference to the superset of persons, the one introduced as generic entity, since the system does not find any hint in the utterance by means of which "a person" could be interpreted as belonging either to the class of consumers or to the class of producers:
[A product is defective when it does not provide the safety which a person is entitled to expect, taking all circumstances into account.]

fact(id116, expect, [experiencer:id22, theme_unaff:id111], 1, tes(fin1_dd10), univ)

fact(id118, entitle, [theme:id22, prop:id116], 1, tes(f6_dd10), univ)

Finally, in utterance 17, a new set of "person" is introduced which "imports a product" and according to the interpretation assigned by the utterance to this set it is included in the superset of the class of producers:

[Without prejudice to the liability of the producer any person who imports into the community a product for sale, hire or any form of distribution in the course of his business, shall be deemed to be a producer within the meaning of this directive.]

class(infon631, id132)
in(infon633, id25, id132)
fact(infon634, inst_of, [ind:id132, class:man], 1, univ, univ)
fact(infon635, isa, [ind:id132, class:person], 1, univ, univ)
fact(id138, import, [agent:id132, theme_aff:id7], 1, tes(f3_dd17), univ)

8. Discourse structure

Narrative has a clear temporal sequence which constitutes its backbone: legal texts lack a temporal timeline and there is neither a temporal progression nor a fixed spatiotemporal location to which events are anchored. In fact, legal texts have no specific temporal reference.

In narrative texts, an incoming clause is included within the previous discourse structure is depending on the following parameters:
- there is no change in the temporal interval, i.e. no new time focus is introduced;
- there is no change in the topic structure, i.e. no new participant is introduced which amounts to saying that either the Topic Hierarchy has no Expected Topic, but a Main Topic, and the state of the discourse is Continue; or, technically a change might be
caused by a Resume, i.e. by the resumption of some previously asserted Topic, from the Discourse Model.

When neither of these conditions are met, the same Discourse Segment is continued either by a Same_Level move or by a Down move (see Polanyi; Webber). However, legal texts lack reference to a spatiotemporal location and no Time Focus is provided, in the sense that there is no fixed external and extensional spatiotemporal location to which events or rather states and entities are bound to. On the contrary, topics hierarchy is still an important tool to organize legal texts even though it might work differently from the way in which it works in narrative texts. These differences are essentially due to the use of referential expressions which might be nominalizations or deverbal nouns which might require recovering the event, activity or state referred to by the base verb; dejectival nouns might require making reference to adjectives used in copulative constructions. Finally, general or generic nouns might either refer to some generic entity or to a specific class which is usually included in the general noun, at certain conditions.

d11. [a product is defective when it does not provide the safety which a person is entitled to expect, taking all circumstances into account.]

same_level:level(11-14)
clause:12-15
topics:[main:id21:person, second:id25:product]
main_fact:defective([id25:product], 1, id44)
ref_int:tint(tr(f3_01), [tr(f3_01), tr(f3_10), tr(f3_10)])
temp_rel:during(tr(f3_10), tr(f3_01))
disc_rel:definition
disc_str:12-[13, 14, 15]
disc_dom:objective
p_o_view:legislator

same_level:level(12-15)
clause:12-16
topics:[main:id21:person, second:id25:product]
main_fact:provide([id25:part, id72:safety], 0, id44)
ref_int:tint(tr(f3_01), [tr(f3_01), tr(f3_10), tr(f3_10), tr(f7_10)])
temp_rel:overlap(tr(f7_10), tr(f3_01))
9. A Legal Text

Council directive of July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products.
The council of the European Communities has adopted this directive.
Having regard to the proposal from the Commission.
Whereas approximation of the laws of the Member States concerning the liability of the producer for damage caused by the defectiveness of his products is necessary because the existing divergences may entail a differing degree of protection of the consumer against damage caused by a defective product to his health or property.
Whereas liability without fault should apply only to movables which have been industrially produced.
Whereas protection of the consumer requires that all the producers involved in the production process should be made liable in so far as their finished product, component part or any raw material supplied by them was defective.
Whereas, to the extent that liability for nuclear injury or damage is already covered in all member states by adequate special rules, it has been possible to exclude damage of this type from the scope of this directive.
Whereas, in situations where several persons are liable for the same damage, the protection of the consumer requires that the injured person should be able to claim full compensation for the damage from any one of them.
Producer means the manufacturer of a finished product, the producer of any raw material or the manufacturer of a component part and any person who, by putting his name, trade mark or other distinguishing feature on the product presents himself as its producer.
The injured person shall be required to prove the damage, the defect and the causal relationship between defect and damage.
This directive shall not apply to injury or damage arising from nuclear accidents and covered by international conventions ratified by the Member States.
Where, as a result of the provisions of this directive, two or more persons are liable for the same damage.
They shall be liable jointly and severally, without prejudice to the provisions of national law concerning the rights of contribution or recourse.
A product is defective when it does not provide the safety which a person is entitled to expect, taking all circumstances into account.
The liability of the producer arising from this directive may not, in relation to the injured person, be limited or excluded by a provision limiting his liability or exempting him from liability.
 Without prejudice to the liability of the producer any person who imports into the community a product for sale, hire or any form of distribution in the course of his business shall be deemed to be a producer within the meaning of this directive.
Whereas to protect the physical well-being and property of the consumer the defectiveness of the product should be determined by reference, not to its fitness for use, but to the lack of the safety which the public at large is entitled to expect.
The liability of the producer may be reduced or disallowed when, having regard to all the circumstances, the damage is caused both by a defect in the product and by the fault of the injured person or any person for whom the injured person is responsible. However, if within three months of receiving the said information the commission does not advise the Member State concerned that it intends submitting such a proposal to the council, the Member State may take the proposed measure immediately.

Any member state may provide that a producer's total liability for damage resulting from a death or personal injury and caused by identical items with the same defect shall be limited to an amount which may not be less than 70 million ecu.
References


