Legal knowledge elicitation: from textual databases to expert systems

Claude Thomasset, François Blanchard[1] et Louis-Claude Paquin[2]

Université du Québec à Montréal[3]

Paper proposed to DEXA '90

International Conference on Data Base and Expert Systems Applications

Technical University of Vienna, August 29-31 1900

Summary

In this paper we intend to present the problems we encounter because of the textual nature of the knowledge in building an expert system for juridical advices on Québec Housing Law. During the knowledge engineering process, we propose to take into consideration both the very textual aspects (morphology, syntax, specific discursivity legal for instance) of the material and the reader's expertise, the jurist for instance. Some simple methodological steps of textual data analysis by computer could help us solve our problem of satisfactory knowledge elicitation.

INTRODUCTION

The aims of the LOGE-EXPERT[4] project are twofold: first, to build a prototype of an Expert System on Québec Housing Law oriented towards non-expert users and second, to produce an ongoing evaluation of the effects of legal knowledge computerization on the legal system and on the users of legal services. This paper presents the latest state of the first aspect of the project. After two versions of a micro-prototype to familiarized ourself both to expert system technology and the legal knowledge engineering[5] we work on an extensive prototype. Lately we settled the general design of the man-machine communication[6]. We now addresses the problems related to the legal text analysis and its knowledge elicitation.

In LOGE-EXPERT, we intend to feed the legal knowledge base with a tribunal decisions data bases. That data bases should intervene at different stages in the building of the knowledge base. At the input stage, the data bank will help its conceptors to validate rules according to decisions given by tribunals specialized in Housing Law. At the output stage, it will illustrate situations corresponding to the user's requests, by selecting pertinent decisions among those included in the tribunal decisions data bank. Meanwhile, it will fill two other functions. It should keep LOGE-EXPERT knowledge base constantly updated with informations taken from the data bank, in two ways: in adding new informations to give the last decisions on matters included in the knowledge base; in inducing structural changes in the knowledge base, justified by substantial changes occuring in the way new tribunal decisions consider now, our specific legal field.

The process of knowledge elicitation in building an expert system is well documented for expertise of technical nature. It is far less the case for expertise of managerial nature because most of the main source of knowledge lies on texts. This is not so at technical level because during an interview, the technician gives heuristics of more help than the general descriptions of the relating texts. Generally speaking, the management is to be accomplished from a written policy. If this policy is complex, it comes with explanatives documents. When it is governmental management of law, the scope of the textual corpus is broad: law, regulations, decree, tribunal decisions, etc. In our case, the Québec Housing Law is to be read with the lights of the decisions, specially to figure out the proper acception of the terms of the law.

Our contribution aims at the knowledge eliciation from textual material. The initial form of the knowledge consist of words, the last is reconstructed out of context in form of concepts and inference rule. Unfortunately, data doesn't arrive in neat little zones and fields pre-marked with delimiters; in our case it comes as a huge mass of texts. In this paper we examine the following aspects related to the very textual nature of the pertinent material available to knowledge elicitation.

What is specific to a text? to a legal text ?

How to efficiently accede the texts?

The reading as an act of knowledge elicitation

How textual analysis assisted by a computer could help knowledge elicitation?

The scope of the answers we propose to each of those points is related to our pragmatic task: to build efficiently a legal expert system.

What is specific to a text?

A text is both inflexible and unsettled: the text is an entanglement of systems that could take several forms We only find in it what have been written: words and punctuation. A minimal definition of a text could be the following: an ordered set of segment written in a natural language recorded on a material (paper or magnetic). A more substantial definition of text should include and present as such the set of linguistical systems, namely: the voicing system (phonological); the outside world referential system (lexical); the internal structure and forms system (morphological) and the the organization and relationship of words within groups, phrases, clauses and sentence system. The higher levels of the linguistic model are less defined. For instance, the semantical level is thought as a kind of calculus on lexical properties of the words and/or their morpho-syntacitcs position in the segment. It is important to take note that the gradual complexity is for didactical presentation. Under real life conditions we have to deal with the unavoidable entanglement of the systems. For exemple, it is virtually impossible to automatically select between two or more potentially contradictory surface categorization without an throroughout description of the deep structure of the text.

After more than 30 years of research in the filed of automatic natural language processors, J. Sowa, a IBM System Research team member stated that:

> "... the successes of language processors on small domains and their failure on unrestricted domains result from the fundamental nature of language. In particular, a large grammar and dictionary are not sufficient to scale up a small system to an unrestricted natural language processor"[7]

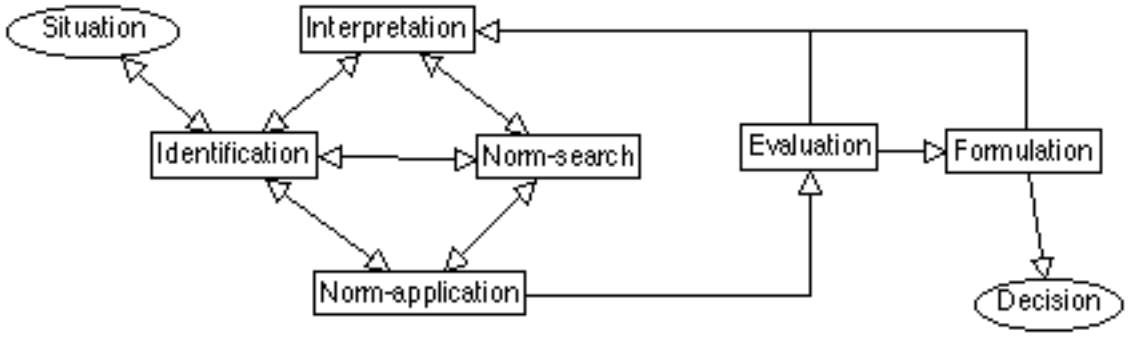
Problems arise when those systemic descriptions intersect higher or at other levels of description we called textual. Among the descriptions of proper textual systems, there is the figure of speech,

the network of argumentation, the communicational environment, etc. Furthermore, not only a minimal knowledge of the particular universe of the text is needed, but also the reader has to be familiar with the social conventions according witch the text had emerged. This dimension of the texts make it to "decode" beyond its linguistic structures.

A text is both inflexible and unsettled. Its life is unpredictable: it could be destroy; it could be duplicated; it could be quoted in an other text with or without indication. This evanescent trajectory is called interdiscursivity and should be take into account. There is several form of text depending on its aim: report, study, directive, free-text answers to opinion pool, interview's retranscription, etc. In the legal domain, the texts have characteristics of their own.

What is specific to a legal text?

We should be aware of some characteristics of the legal texts , characteristics that have been already pointed out by different authors. For instance, both theoricians (computer scientists or logicians that tackled the field of law as a mean to test their research program) and practicing lawyers that had tried to develop a “machine-like” understanding of legal texts have came across several difficulties. For those dealing with statutes, it became apparent that: “the structure of the text of the statute is no longer irrelevant but dictates the nature of the formalization[8]” and that the “definition of a concept” may have been distributed “ across a number of fragments”. Others[9] still, have try to described the interactive play involved in the legal reasoning process.



Overwiew ; different activities involved in the legal reasoning process
From Peter Wahlgren, 1989, Legal reasoning — A Jurisprudential Description, Fig 1, p.150

This schemata, that could perhaps be translated in a set of rules (with some of them expressly recursive), illustrate well the complexity of the process in which a legal text is interpreted and/or produced. But most authors,in trying to described legal texts as a by-product of reason, subscribing to Toulmin[10] interpretation for example, had seems to forget that texts are “artefacts” who lives independently from the individual reason that had given rise to.

Legal texts could be classified according to a norms hierarchy in a written legal system as the Québec legal system is. After the **Constitution**, which is the founding legal document, **Laws**, as documents issued by legislative bodies, are considered as the supreme norms. Then come **regulations** adopted by executive agencies, to give effects to laws. **Courts and Tribunals decisions** apply laws and regulations to specific situations, when litigations occur. Some of these decisions are published and could become cases to be used as **precedent** for other decisions

To illustrate that hierarchy in legal texts, 1) we have chosen excerpts from article 1659 of the Québec Civil code related to repossession of premises for the benefit of landlords. As we have selected an article of the Québec Civil code that usually does not require regulations for its implementation, we cannot find a corresponding regulation in the field of repossession. For example, no regulation has been educated to give a definition to the legal concept "**landlord**". We have then, to investigate through tribunal decisions and legal doctrine to assess all the meanings this concept could carry in the field of landlord and tenant relationship;. 2) Nevertheless, we have selected section 3 of a regulation related to obligatory dispositions of a residential lease, adopted to

implement the *Law creating the Québec Rental Board* (la Régie du logement) *and amending the Civil code,* to illustrate the way that kind of legal text intervene in the legal norms hierarchy. We have completed these legal texts with two types of Tribunal decisions: 3) one from the Québec rental board which has been published and 4) its confirmation by the appellate Tribunal.

How to efficiently accede the texts?

The information is always abundant; to access a specific information, one must go through a huge mass of non-pertinent informations. We only can assimilate a small proportion of the available relevant informations. The management of the information need a computarized information system and a hierarchy of administrative structures. The first computarized information systems were bibliographical: the access to the document was not one of its content but of its localization in a library shelf; the textual information systems are more recent. The browsing of information is a time consuming essential operation. Information retrieval researches are oriented towards the improvement of two measure: recalling rate and precision rate. The recalling rate is the percentage of the relevant documents found against the total number of relevant documents. The precision rate is the percentage of the relevant documents found against the total number of document found.

Every textual information system has its own (specialized) data structure and associate query language that one has to learn before getting access to the information. A format (rows, columns, fields or zone) or an index made in advance cannot foresee all the users requests. It is specially so in the interactive, free-association and browsing process of reading with the help of the computer. Beside the two rates, a textual information system should take input data, in a loose structured formats (ad hoc protocol); it should allow easy, interactive free-association and browsing (hypertext, RDBMS). Last, it should not demand too much work to get data into the system.

There is however databases that are specialized in text handling; lawyers do already know them; they exist as “on line” services: you can have access to them through phone via a modem. They are Lexis or Westlaw in USA and Quicklaw or Soquij in Québec. As such, they do require

from the users the learning of new skills: their query language is proper to each of them, so almost incompatible. For exemple compare " s locateur /5 usufruitier & possession" from QUICKSEARCH (from Quicklaw) to "..repere locateur avec usufruitier" from SOQUIJ. But there are more limitations: they are only available trough a communication process that cannot be easily built in an expert system without taxing the skills or the patience of the users.

One should dispose of a modem, a phone line, a communication program with a familiar(?!) user interface. One should expect delays from datapac and the server for a sign-on procedure. Finally one should be ready to pay, because those on-line databases charge on a time basis for their consultation. Their existence do point out that it is possible to transform legal text into factual data by the process of indexing [for instances Soquij in Montéeal utilize the software STAIRS (©IBM) for that purpose; Sonar for the Mac can almost do the same for a micro]. The result of this process permits the end user to ask for the occurrence(s) or the co-occurrence(s) of word(s) that are linked (more or less obviously) to the topic of his interrogations. This type of interaction with the data take at least a couple of hours for a law student to master, because at the beginning the questions that spring to mind as to be translated in a query language, that is simple but unforgiving in the number of debris it can accumulate, and the search as to focus on the occurrences of words, not notions or concepts. In a sense then they are just a "too good tool" to be build in: they are far to general, cumbersome not to intimidated the end user.

In order to accomplish all the tasks the tribunal decisions data bank should fulfill in relation with LOGE-EXPERT knowledge base, we could not limit the dimensions of the decisions to summaries and key-words. We need the tribunal decisions, in their full texts. The needed selective content access is not only with pattern of words but also by a conceptual pattern. The main problem with conceptual patterns is that an indexation is needed. Indexing is intended as the transformation of proheminent terms from a text into factual data . This tagging process was, up to lately perform by human readers,.now it is made by computer in rough way in a sens but at a relatively low (software and a personnal computer) cost and at high speed when the text are already machine readable. This indexation, named key word in context (KWIC) but could be selective when the

KWIC is the result of an elaborate.pattern matching of word chains. The criterion for the match should also be structural. This could be seen in two ways: the first is linguistic "we need a full structural description as deep as possible (at any cost?)"; the second is more pragmatic in the sens that the text is not by itself an objetct of study.

Our textual database was primarily aimed at collecting all the decisions of an administrative body, " La régie du Logement", and all decisions of superiors court pertaining to those decisions. "La régie du Logement" does published some of those decisions, about 150 per years (which represent less then 1/100 of the decisions that, for the main part, [fixation de loyer] are fairly short and "automatic"). Since we were unable yet to capture the decision at the source, as text file from a word processor, we, instead opted for a mechanical transcription, by means of optical character recognition, of only a small part of the decisions: those relative to "reprise de possession" as the index of the published decisions defines them, about 17 for 1988 for instance. After their transformation into a text file, we subject them to an orthographical revision (for , and ') and save them in two distinct format: _ one that we could manipulate [like inserting TAB to imported them into a database] and _ one that could preserve the formal appearance of the decision so that a consultation could produce an on-screen facsimile of the retrieved decision(s). The corpus is, at first, quite small but due to increase as we expend our coverage in time and in the domain of the law.

The reading as an act of knowledge elicitation

We have seen that the knowledge doesn't lay at the text surface. The only way to overcome the lack of certainty concerning the meaning of a text and still use the computer is to include the reader into the model of the meaning construction process. This intuition is confirmed both by the psychological and sociological latest theory stating that a univocal meaning is not set down in the text but constructed by the reader through his cognitive structures and his socialization. But reading alone prove to be an inconsistent knowledge elicitation device at the low level of recognition; we need the computer regularity in complex pattern matching situation to improve the reader making of

better judgmental choice. Gathering informations is what human perform worst, make inferences; take decision on uncertain facts is what human perform best. In this perspective, the computer accumulates every context needed by the expert-reader to construct a meaning by correlating informations gatthered throughout the corpus. This model is attractive because it includes the implicit possibility of a partial integration of the reading "expertise" (with the knowledge engineering inquiry techniques?).

According to J. Wroblewski, legal language comes from natural language[11] and adds to it, specialized words and specific meanings, corresponding to the legal nature of that discourse. Even if some researchers try to establish that a legal grammar specific to legal language(Araj Houda)could be established, for the purpose of this paper, we may accept the proposition of J. Wroblewski. The difference between natural language and legal language would not syntactic but semantic, and depends on the words as well as their specific meanings[12]. He distinguishes legal language, the language of legislative bodies, from meta legal 33discourses, among which he classifies the discourse of Courts and tribunals, the discourse of legal authors (doctrine and jurisprudence), and the commun legal discourse used by lawyers in their oral arguments in Courts or by the general population when they talk about law. The legal documents we use in LOGE-EXPERT knowledge base, belong to all these kinds of legal language and discourses. But, the tribunal decisions data bank is devoted only to tribunal decisions which are expressed in a meta legal discourse, the language in which the law is applied by judges, to specific situations.

If we analyse the way a legal expert used to read each kind of these legal documents, we may establish patterns of reading operating in order to catch all their possible meanings. These patterns of reading could be illustrated by successive readings of a legal document, each reading adding a new level of understanding and dessiffering the specific meaning of legal concepts it includes. Depending on the legal expert's skill, these levels of reading could be accomplshed in one or several reading experiments. The **first level** of reading is a common sense reading, or syntagmatic reading, that means we read it as if it is expressed in natural language. A **second level** of reading or paradigmatic reading, implies that we focus on words as legal concepts. The **third**

level of reading will be done to compare the legal document to a specific situation. The **fourth level** of reading will try to give to the legal document meanings, in order to apply it to a specific situation by interpreting it according ad hoc methodology in use in the legal universe. A fifth level of reading could be reached when we want to analyse the symbolic dimensions of legal document. We have experimented this pattern, with our samples of legal documents, we have previously selected.

Article 1659 of Québec Code civil:

> Le locateur d'un logement
> peut
> en reprendre possession
> pour s'y loger
> ou
> y loger
> ses ascendants
> ou
> descendants
> son gendre, sa bru, son beau-père, sa belle-mère, son beau-fils,sa belle-fille
> ou
> tout autre parent
> dont il est le principal soutien.

At the **first reading level** (syntagmatic), we scrutenize the meanings of the words in the context: At the **second reading level** (paradigmatic) we focus on legal concepts, such as: *locateur, ascendant, descendant, parent, principal soutien, logement***.** To better understand the meanings of these legal concepts, we have to get out from the document, in order to other possible meanings in co-texts. At that stage, the legal expert reader looks out for indications in other parts of the legal document (law or regulation), or in other decisions or cases, the legal cases discourse along J. Wroblewsky classification, as well as among legal doctrinal works, designated by J. Wroblewsky as scientific legal discourse. We did that search to get all the meanings of the legal concept *locateur,* and we could duplicate it for all the other legal concepts identifed in that segment of section 1659 of Qubec Civil code. The next step is the **third reading level** or the "analyse de correspondances", which usually is dedicated to apply the legal document as understood after the second reading enriched with the search of meanings,to a specific situation. At that stage, the legal expert used to qualify facts and makes hypothesis about the applicability of that legal documents. When

necessary, a **fourth reading level** could be done, in order to give effect to legal concepts to the specific situation under consideration. At that stage, a legal expert reader will focus then, on general legal principals such as "the legislator 's will" or "the general meaning of the law", in order to enlarge the legal concepts meanings obtained from the precedent searches. Usually, for the implementation of that segment of article 1659 of Québec Civil code, the legal expert will invoque the property rights of the landlord and its components like "usus", "abusus", and "fructus", or the tenant right to stay in the dwellings according to article 1657 of the same Code. We are very near at that stage to the next level of reading. A **fifth reading level** is still possible, in order to identify the symbolic dimensions of a legal documents. This step is occasionnally done when the legal expert wants to give effect to general assertions included in segments of that texts. But for the purposes of the analysis of article 1659, it is not necessary to expend the level of reading to that stage.

Article 3 (Règlement sur les mentions obligatoires du bail, de l'écrit et de certains avis prévus au Code civil, R.R.Q. R-8.1, r.2)

le bail
s'il est
écrit
doit
contenir
    la désignation du logement
    le montant du loyer
et
    la date
        du début
        et
        de la fin
        du bail
s'il est
à durée fixe

This is the segmentation done at **syntagmatic level**; here are the selected legal concepts at **paradigmatic level**: *bail, loyer, durée fixe, logement*. At this stage, it is necessary to find in the habilitating law, the meanings of these legal concepts. For example, we know that in Québec Civil code, «un bail à durée fixe» implies the opposit concept «bail à durée indéterminée», and «bail écrit» implies the opposit concept «bail verbal». If the habilitating law could not help to understand these meanings, we get new legal sources to explicit them. The **third reading** is the same kind of reading than for laws, but instead of applying law, we want to give effect to a regulation to a specific

situation. Facts are systematized to fit with the regulation and its habilitating law. At that stage of the **fourth reading**, it is important to evaluate the suppletive or imperative nature of the regulation and to make explicit the intention of the executive body which has enacted it or the legislative body which has delegated these powers. Even in regulation, we may find symbolic wordings, which could have effect without any means of implementation; this is considered as a **fifth reading**.

Germaine Lalonde-Sarault et al. c. Huguette Masse, Gustave Hébert, régisseur, Régie du logement, Verdun, 34-870115-017-G, le 3 juin 1987. J.L. 88-12
Reprise de possession

This is an excerpt from a published decision. We have selected the final part of the decision, because it sums up the legal reasoning of the administrative judge in that specific case. We have introduced the excerpt with key words enumerated for the purpose of editing the decision to publish it in Jurisprudence-Logement, a case report under the responsibility of the Québec Rental Board A legal expert will read quite differently a law or a regulation, and a Court or tribunal decision. We may rely this practice to the classification set up by J. Wroblewsky. The formers are legal language according to that classification, the latters are legal discourses. Nevertheless, we may establish a pattern which has some common elements. The first stage, when a decision is edited to be published, is to read key-words, and abstract if it exists. It does not in our example. Then a legal expert will go the conclusive part of the decision to look at its motivation. The second stage will depends on the reader's purpose. He could be satisfied with the **first reading** if he only wants to update his knowledge about that legal field. But if he intends to use this decision in an argumentation along or against that decision, then his next readings will try to analyse the decision in order to give it the meanings he needs to complete his demonstration.

The **second reading** will focus in analysing the facts of the case, in order to compare them with the specific situation he is confronted with. We may include into facts, all the procedural steps to get to the tribunal. The **third reading** will pay attention to the argumentation of each part and to the way the administrative judge receives them. At this stage, the legal reasoning followed by the judge will be thouroughly analysed. The **fourth reading** will help to evaluate the applicability of the decision to other situations. The legal expert at this stage will get off the decision and compare

the situation under consideration with the situation sanctionned in the decision. The **fifth reading** will mostly interest legal academics who try to analyse the effectiveness of laws and regulations through Courts and Tribunals decisions. At that stage, it is important to analyse the work of the judge in expliciting the meanings of legal concepts and giving them consequences. Legal concepts become objects of debate. In our case, we may confronte key-words with the conclusions of the administrative judge: we understand that the issue is to know if the lessor (locateur) who is «usufruitière»of the rented premisses, is entitled to get the repossession for the benefit of her brother, who is «nu-propriétaire». The conclusion is no, because the «nu-propriétaire» has not the right to use premisses during the life time of the «usufruitier». The article 1659 of Québec Civil code could not apply to the situation submitted to the judge.

There has been an appeal from the part against whom the decision has been issued to the appelate tribunal. The latter has confirmed the first level decision. The judgement has been also published.

> **Germaine Lalonde Sarault et al. c. Huguette Masse, monsieur le juge Jean-Louis Lamoureux, Cour provinciale, Montréal, 500-02-022192-871, le 29 janvier 1988 , J.L. 88-42.**

The same pattern of reading could be followed to analyse this appelate decision. This decision worth special interest because the appelate judge has opposed to the right of lessor to the repossession of rented dwellings, the tenant's right to stay in the rented premisses. He gives effect to that right in interpretating restrictivly the lessor's right, and the legal concept "lessor". At the difference of a published decision which is edited and is processed with key-words and abstracts, a non-published decision is a legal document which requires from its readers that they did themselves the editing process to get key-words and a summary of the essential features. That means that the reading pattern has to take account of this processing steps if we want to simulate the reader's work. The other stages of a legal expert reading pattern are similar in both cases.

How texual analysis assisted by a computer could help knowledge elicitation?

What a computer could do to simulate legal expert reading patterns? We have established two reading patterns, according to types of legal documents a legal expert is confronted with. If the legal document is law or regulation, the reading pattern includes the analyse of contexts and co-texts of the written document in order to establish the specific meanings of words expressed in a legal language. If the legal document is a court or tribunal decision, we are confronted with a multi-dimensional analysis of that kind of legal discourse. The reading pattern has to cope with the argumentative nature of a decision, which implies that the judge has to make explicit the reasons which justify his decision. The legal expert usually find all the references of the judgement motivation in the decision. He has not to search for them because the adversarial procedure facilitates the expression of the pro and contra arguments during the proceedings. The judge has to consider them and to make his own opinion in order to write his decision. The contexts and co-texts are included in the decision which give it its multi-dimensions size.

For our purpose, which consists in building a data bank with tribunal decisions in Housing Law, the computer has not to procede through all stages We only need a first and a second levels of reading in order to identify legal concepts in use in Housing law domain and common sense words which characterize specific situations. That means that the automated text analyses should be able to identifie common sense words and legal concepts. It is not necessary to require that a computer accomplishes the other levels of reading of legal texts.and discourses The human legal expert will do it at the input stage and the human legal expert users will intervene at the output stage.

The knowledge we are looking for is to be represented in a well defined data model such as networks, frames, valuated objects, inferences rules, etc. At the other end, a term comes always included in the near context of a nominal group where it gets specification and determination from others terms or from various qualities. Between those two poles is the human expert reading. The relative slowness of the human process and the potential lack of attention make the computer's assistance desirable. Textual analysis with both its linguistic and discourse sides offers a suitable theoritical framework. Our aim is not the exhaustive description of all the textual systems, but the dissociation of the information from it's enunciation conditions. Because the highly structured

approach are generally static, we favour the gathering of analysis procedures from various, sometimes opposite, point of views. But the scope of these procedures should be cautiously tuned to methodological principles.

The control of these analysis should not be let to the automat. By taking the exclusive point of view of the author's (legislator for instance) intentions, the analysis will be normative. To avoid this bias, the reader must have the control over the sequence and the modulation of textual manipulations and mesurements. Prior to any analysis, the reader must have formalize an hypothetical model of its corpus. The model should cover both formal and content aspects. A description in term of break and consistency in front of the natural language should be made. Every significative difference at any systems should be taken down and investigated. From this model, the parsing focus and needed adaptations together with the "grain" of the pattern matching will show up. The reader should have the proper cognitive structure in order to match statements to events happening to "real word" objects. The model will help 1) formalize the problem's space; 2) evaluate the better computationnal approach to produce the needed description at the suited depth; 3) fix boundaries to the corpus, withch texts fit in and why thoses and not others? 4) determine the best description needed both by the model and the aims of the project, knowledge elicitation for instance.

The textual description is made in terms of groups, properties and relations. The more static element is the lexicon; its categorization out of context could be make automatically. If one want to search for surface structures, a morphological description is a minimum. The resulting informations should be tied to the lexical form; the valuated object seem to be the more adequate representation: beside the morphological one, a slot must be devouted to every aspects the analysis project is monitoring. This slot is to be filled when someting signifiant is observed or inherit from observation at other levels. Looking for reference of concepts, the nominal group interest us. Its variations are many but most of them are localized in a small word set. A surface pattern with masks and inform by a projected morphological description seem sufficiant to give interesting results. In a computationnel framework, the conceptual indexation could be repalce by nets of associated concepts; one node of the net being tagged with the concept name. The numerical co-occurence of

two words in the same limited context represent a statistical validation of their expressed relationships within the text. The KWIC search with annotation facilities represent an alternative way to interact with textual data: to sort and search in a multi-dimensions (proprieties).space.

On a text whose words have been morphologically categorized, vocabulary control could be peformed: 1) tie the words forming idiomatic expressions to track terminological phrases from the co-occurrence of categories, e.g. *traitement de textes* {[noun] preposition [noun]}; 2) reduce the different surface forms of the words to their dictionnary entry (beside the verbs, in french each adjectives takes four different forms and the noun two), this task is called lemmatization; 3) cross-reference the synonymy, one term has to be prefered and the others non-prefered the links have to be validated by experts; 4) restitute the incomplete wordings with correct interpolations, substitution of the anaphora (pronom used as substitute); this last operation is up to now out of reach of the algorithmic modelisation, because it rest on jugment.

On a controlled vocabulary, a concept dictionnary has to be built. A restriction of the lexicon to the words categorized «noun» is then to be reduced manually by experts in order to keep only those concepts related to the field of expertise. Then a transformation of the terms into concepts is performed. It is an operation of abstraction; a word or a locution is taken out of the context and the contextual relavant information is annotated. It is also an operation of condensation; for each concept, a general model of all the contexts found in the corpus is develop by classification of the dependant terms and adjectives. This model could then be implemented into an index for the information retreival; into frames of an expert system, or into nodes of an hypertext. Every nouns validated by domain experts has to be analyse in terms of configurations (called ingredients) that are associated with these concepts are seeked out. The search for hierarcherical relations explicitly stated is then made in term of word patterns such as « x has a y», «x is made of y», «x is part of», etc. Adjectival forms found in located contexts reveal quantifiers and argument scales that virtually position other possible qualitative or quantitative value. On the other hand, analyzing verbal groups helps in the inference rule writing process. Indeed, the study of action verbs allows the tracking of

object-defined operations. Their inflection and context provide thus modulation (active, passive, necessary, optional, etc.), localization and temporality.

CONCLUSION

Textual analysis seems to be the available best approach to knowledge transfer from texts to expert systems assisted by computer. We advocate the use of computer packages for processing the search of knowledge. Resting on morpho-syntactical structural selective pattern matching, we use the metalanguage inherent to the text itself to single out organized and hierarchized invariants by their recurrence. It meets criteria of consistency, objectivity, reproductiveness and independence as to the problematics defined in the texts. Furthermore, a methodological approach with general purpose analytical tools, seems to us both more trustfull and more helpfull than a "black box" application.. We believe the main obstacle to a full use of texts in expertise transfer lies in a lack of understanding of the nature of text. The words that compose it do not necessarily refer to realty through concepts, but may serve to recategorize them. Gaps and modification in the referential structure of the text bring the reader to produce several inferences. Finally we believe that experts-reader, jurists, should be include at all the steps of software or methodological developpements.

---

[1]Groupe de Rercherche en Informatique et Droit (GRID)

[2]Centre d'Analyse de Textes par Ordinateur (ATO)

[3]Université du Québec à Montréal
Case postale 8888, succursale "A"
Montréal (Québec) H3C 3P8

[4] This project, witch runs since January 1987, received an award from the Donner foundation. The other participants are (René Hébert).
Since now, a few contribution are to be published :

[5]THOMASSET, C., BLANCHARD, F., HEBERT, R. La formalisation du savoir juridique et la construction d'un prototype de système expert en droit . Communication présentée à la IVe Conférence Canadienne sur l'informatique et les processus d'enseignement du droit. Université Laval, juin 1988, 22p.
THOMASSET, C. "Expert System and Legal Formalization: Evaluation of a prototype in Québec Housing Law". Paper presented at the International Conference on "Intelligence and Society", International Social Science Council, European Coordination Centre for Research and Documentation in Social Sciences, Vienna, March 1988, to be published by Reidel (Netherlands).
THOMASSET, C., " Expert System in Québec Housing Law: from HOME-Expert I to HOME-Expert II", paper presented at the International Conference on Law and Artificial Intelligence: Expert Systems in Law. CIRFID, Bologna, May 3-5, 1989, 29p.

[6]THOMASSET, C. PAQUIN L. C. "Expert Systems in Law and the Representation of Legal Knowledge: Can We Isolate It from the Why and the Who?", Proceedings of the Third International Congress on: Logica, Informatica, Diritto: Legal Experts Systems, Florence, 1989, Istituto per la documentazione giuridica, vol. 1, pp. 751-771.

[7] SOWA, J. F. "Multi-Domain Semantic Theory" draft given by the author at a Montreal conference dated November 28, 1988.
See also: SOWA, J. F."There's More to Logic than the Predicate Calculus" draft given by the author at a Montreal conference dated November 28, 1988.

[8] Routen, Tom. Hierchically Organised Formalisation. The Second ICAIL. Vancouver, BC: ACM, 1989. p. 242-250.

[9]Wahlgren, Peter. Legal Reasoning _ A jurisprudential Description. The Second ICAIL. Vancouver, BC: ACM, 1989. p. 147-156.

[10]Toulmin, Stephen. The uses of argument, C.U.P. 1958.

[11]Wroblewski, Jerzy. "Les langages juridiques: une typologie" Droit et société ****,p.14

[12] Ibidem, p. 15

Article 1659 of Québec Code civil:

> *Le locateur d'un logement peut en reprendre possession pour s'y loger ou pour y loger ses ascendants ou descendants, son gendre, sa bru, son beau-père, sa belle-mère, son beau-fils, sa belle-fille, ou tout autre parent dont il est le principal soutien.*

Article 3 (Règlement sur les mentions obligatoires du bail, de l'écrit et de certains avis prévus au Code civil, R.R.Q. R-8.1, r.2)

> *Le bail, s'il est est écrit, doit contenir la désignation du logement, le montant du loyer, et la date du début et de la fin du bail, s'il est à durée fixe.*

Germaine Lalonde-Sarault et al. c. Huguette Masse, Gustave Hébert, régisseur, Régie du logement, Verdun, 34-870115-017-G, le 3 juin 1987. J.L. 88-12

> *Considérant que le frère de la locatrice, n'est ni locateur, ni propriétaire du logement;*
>
> *Considérant que « la nue-propriété contient le droit de jouir pour l'époque où s'éteindra l'usufruit actuellement existant», le droit pour les nus-propriétaires indivis de reprendre possession du logement n'est que différé;*
>
> *Considérant la preuve;*
>
> *Rejette la demande de la locatrice.*

Germaine Lalonde Sarault et al. c. Huguette Masse, monsieur le juge Jean-Louis Lamoureux, Cour provinciale, Montréal, 500-02-022192-871, le 29 janvier 1988 , J.L. 88-42.

> *Pour les motifs exprimés dans la décision du régisseur, le Tribunal en vient aux mêmes conclusions que le régisseur.*
> *Le principe énoncé à l'article 1657 du Code civil que le locataire a droit au maintien dans les lieux exige que les articles 136.1 de la Loi sur la Régie du logement et l'article 1659 du Code civil soient interprétés restrictivement.*
>
> *L'appelant gaston lalonde n'est que le nu-propriétaire de l'immeuble. Il n'en a ni «l'usus» ni le «fructus». De plus, il n'est pas le locateur au sens de l'article 1659C.c. C'est Germaine Lalonde Sarault qui est le locateur non seulement apparent mais réel du logement dont on veut reprendre possession. C'est elle qui signe les baux, retire les loyers et voit à l'adminsitration de l'immeuble dont elle a l'usufruit. Ce n'est qu'occasionnellement qu'e;lle se fait aider par son frère, le co-appelant.*
>
> *Par ces motifs, le tribunal:*
> *Rejette l'appel sans frais.*