Loge-Expert[*]*:*

Strategies to Integrate Legal Knowledge Modelization, Non-Expert User Interface, and Textual Data Base into the Developpement of an Expert System in Law

Claude Thomasset, GRID,

François Blanchard, GRID,

Louis-Claude Paquin, ATO

Université du Québec à Montréal[°]

Short title: Loge-Expert, An Expert System in Law

---

[*] Loge-Expert is part of a research funded by the Canadian Donner Foundation.

[°] Université du Québec à Montréal, C.P. 8888 Succ. A, Montréal, Canada, H3C 3P8;
Tel.: (514) 987-7932; Fax: (514)-987-4784

Loge-Expert:

Strategies to Integrate Legal Knowledge Modelization, Non-Expert User Interface, and Textual Data Base into the Developpement of an Expert System in Law

Abstract:

Our paper will present three problems we have been exploring for the last two years. The first concerns modelization and formalization of legal knowledge. The second one refers to the communication between the user and the machine, since we have chosen to develop an expert system in Housing law oriented towards the general population. The third problem relates to the textual data base we have built in order to document the knowledge base of Loge-expert. At the moment, it contains some of the textual legal sources traditionally used by lawyers.

From this two-year experiment with Loge-expert, we have learned that the development of an expert system in law is an iterative operation which includes simultaneous exploration of different dimensions such as the conceptualization of the specific legal field as well as the conceptualization of the communication devices between that legal expertise and the user or the conceptualization of the linguistic and textual dimensions of that legal expertise.

Key words:

Law / Knowledge Elicitation / Modelization / Formalization / Interface / Non-Expert User / Textual Data Base /

Introduction

The conception and the development of an expert system are an iterative rather than a linear operation. We have learned from our experience that they resemble an obstacle course that we must negotiate to cross the finish line. Difficulties came not only from the non-existence of a methodology which would cover the entire process of development of an expert system and from the inadequacy of a specific methodology for some types of expert systems, but also from the inability to adequately define the field of expertise or to modelize the problem to be solved.

Our paper will present three problems we have been exploring for the last two years and which we succeeded in conceptualizing in order to clarify our progress. Our progress was slower than expected because of administrative problems as well as the unavoidable time required to get each phase of our development process ready before going to the next. Our trial and error working strategy has led us, through successive loops, to the results we are now ready to disclose.

As a preliminary point, we will make a brief presentation of the aims and the initial choices we established before undertaking the development of Loge-expert. In subsequent points we will describe three problems we have faced. We will conclude by stressing the important role played by legal experts in the process of formalization of legal knowledge.

Since the very beginning of Loge-expert, we have been confronted with a dilemma: A.I. developments are still at a level of fundamental and even sometimes highly speculative research; nevertheless, some of its applications are on the way to being implemented in the real world. Loge-expert is an example of this dual process. On one hand, it should be an aid to a specific legal decision-making process, available to the general population. On the other hand, it represents an opportunity to proceed to a thorough evaluation of the application of A.I. to law. Two questions sum up this controversial situation: is

research in A.I. of some help in building an expert system for the dissemination of legal knowledge among the general population? What are the limits of A.I. applied to law, from a technological as well as from a socio-legal point of view?

In order to share this exciting experiment and this feeling of creative anticipation, we have published each stage we have gone through in this exacting process of legal knowledge formalization in different papers (Thomasset, Hébert, 1988; Thomasset, Blanchard, Hébert, 1988; Thomasset, 1988; Thomasset, 1989; Thomasset, Paquin, 1989; Thomasset, Blanchard, Paquin, 1990).

At the outset, we made two decisions from a pragmatic point of view which have shaped Loge-expert. We used the local resources available at our University: an expertise in housing law and D_Expert, a generator of expert systems developed at ATO research laboratory.

Another consideration which has been taken into account in selecting the legal expertise to be formalized is that housing legal knowledge should be rendered accessible to the general population. This social goal is based on Montréal statistics on the predominantly tenant proportion of the city's population (78% according to the 1986 Canadian Census) and the high number of housing actions which are adjudicated each year by the Régie du logement, the most important Québec tribunal according to its case load (70,000 to 90,000 per year).

As for the choice of D-Expert, we wished to benefit from the opportunity to work closely with its developer in order to get it tailored to the constraints of legal knowledge formalization. D_Expert offers a knowledge representation structure based on valuated objects under the hierarchy of "granules", features and values (Paquin, 1987a, 1987b). A "granule" corresponds to a concept which may include a string of other concepts that we will reduce in turn to "granules" aggregated by means of links as well as inference rules. Each granule is specified by features which are characterized

by values expressed in terms of numbers or symbols.

Having completed our contextual presentation, we shall now present three of the problems we have analyzed.

The first concerns modelization and formalization of legal knowledge. The second one refers to the communication between the user and the machine, since we have chosen to develop an expert system in Housing law oriented towards the general population. The third problem relates to the textual data base we have built in order to document the knowledge base of Loge-expert. At the moment, it contains some of the textual legal sources traditionally used by lawyers. At a later stage we intend to integrate the modelization of strategies adopted by lawyers in the selection of relevant information from these textual legal sources with the data base in order to proceed through their problem solving process. We are now working on methods of interviewing legal experts to extract the research strategies they employ when they are engaged in a legal problem-solving process.

1. The modelization and the formalization of legal knowledge

This stage was the most demanding since we were faced on the one hand with the learning process of the cognitive structures proposed by D_Expert, and on the other hand, with their application to our legal field in order to represent it as inferential rules. From it we have drawn some conclusions which we will later compare to similar experiments of developments of legal expert systems in subsequent works.

1.1. Macro-modelization

Québec Housing Law is a rather new legal field which has grown for the last twenty years with the adoption of legislation and regulations in accordance with housing policies enacted by federal, provincial or municipal bodies. The selection of specific questions from this heterogeneous field proved to be difficult in the absence of

consensus in legal doctrine and tribunal decisions. Finally, the residential lease was retained with three specific questions: the payment of rent by the tenant, the repossession of a dwelling by a landlord, and the decrease of the rent in relation to the landlord's default to fulfill his obligations. The latter question will be formalized in a further stage of our works.

In our first attempt to formalize the expertise required for the payment of the rent by the tenant, we were confronted with the necessary modelization of much more legal knowledge than that specifically related to the question. We had to take comprehensive legal concepts into account which were broader than those we were working with to be able to write correct inference rules.

In order to modelize the rent payment question, we had to include key concepts belonging to the general theory of obligations in the knowledge base of Loge-Expert. We named this operation macro-modelization (Thomasset, Paquin, 1989). Figure 1. illustrates our modelization strategy to delimit our specific legal field. Even if the residential lease is an exception to the key principles of the general theory of obligations, we had to formalize them in order to get correct inference rules.

[Insert Figure 1 about here]

Legal reasoning in a specific field is not limited to the use of its specific concepts. It refers to concepts belonging to larger legal domains if they are required to get correct inferences about that specific field. Legal knowledge looks as if it were organized in hierarchical structures. Thus reasoning at a lower level such as the residential lease, implies reference to concepts related to a broader level, such as the general theory of obligations. Fortunately, D_Expert did not put constraints on the sequences along with

developing the knowledge base of Loge-expert. We were able to complete the modelization of those general legal concepts required after having modelized the more specific ones.

1.2 Micro-modelization

Our modelization strategy results in a dictionary of concepts belonging to the residential lease as well as to the general theory of obligations such as the capacity to contract, mutual consent, the object of the contract and the cause of the contract. When we arrived at the modelization of the specific question of repossession of a dwelling by the landlord, we faced a difficulty of a different nature. Instead of looking for more comprehensive legal concepts, we had, on the contrary, to specify the meaning of some legal concepts in a very restrictive way. For example, the concept “locateur” (landlord), which offers no difficulty in most situations where a residential lease intervenes, is a problematic concept in situations related to repossession, since no definition is provided in either the *Civil Code* or in the *Loi sur la Régie du Logement* and its corresponding regulations. We had then to try to get its specific meanings through legal doctrine and tribunal decisions.

We had to establish a complete list of all the possible meanings of the concept (owner, co-owner, usufructuary...) for the dictionary of concepts, even if only some of them were relevant to repossession. This operation was called micro-modelization (Thomasset, Paquin, 1989).

As opposed to the macro-modelization operation, which consists of locating relevant concepts from more comprehensive legal knowledge levels, micro-modelization draws us in the opposite direction, towards more and more specific levels of legal knowledge. For example, the concept “locateur” (landlord) was scrutinized in order to establish the complete list of all the meanings it might possess in the residential lease field. This enabled us to select those which are relevant to the specific question of repossession. At

that point we reached the bottom end of the possible meanings of the concept "locateur". We could not further brake it down into a smaller unit.

This micro-modelization operation which results in an identification of all the possible meanings of a specific concept, is completed with the establishment of links between those different meanings. For example, meanings such as mandatary, which sometimes qualifies the concept "locateur", should be regrouped into the category of landlords ineligible to repossess, while other meanings such as "owner" or "usufructuary", which could sometimes define the same concept, should belong to the category of landlords eligible to repossess. On the other hand, the meaning "purchaser" could belong to either category according to its non-convergent interpretation by legal doctrine or tribunal decisions. We can complete the dictionary of concepts at any stage of the legal knowledge formalization in the same manner as we have done in the macro-modelization operation.

1.3 the Reduction -Standardization

With the lessons we learned through the macro and micro-modelization operations, we developed a knowledge base organized in a tri-level structure of "granules"-features-values, which includes legal concepts as well as non-legal ones. The purpose of this modelization at the moment is to formalize a legal expertise about the specific question of repossession. This legal expertise results from legal knowledge obtained through different legal sources such as legal provisions and statutes, regulations, tribunal decisions, legal doctrine and legal expert know-how which are activated by models of legal reasoning. Figure 2 sums up those legal sources of our expertise.

[Insert Figure 2 about here]

From these legal sources of expertise we extracted the concepts to be formalized in order to write the relevant inference rules. For example, with the repossession question we selected legal concepts such as “locateur” (landlord), “locateur eligible” (eligible landlord), “bénéficiaire de la reprise de possession” (beneficiary of the repossession), “logement” (dwelling), “loyer” (rent), “bonne foi” (good faith), as well as non-legal concepts such as “type de logements” (type of dwellings), “logement disponible” (available dwelling) and “localisation” (localization). These concepts are opposed to facts which should also be modelized since Loge-expert is supposed to give access to the general population who are invited to submit their conflictual situation to the formalized legal expertise. This factual data must be correctedly interpreted in order to activate relevant inference rules. The knowledge base should consequently include legal and non-legal knowledge as well as factual data. The modelization of these different objects leads to standardization. These concepts and facts are classified in accordance with an inclusive strategy in order to facilitate recognition of the broadest possible conflictual situations. This strategy results in a limited choice of typical conflictual situations, which eliminates the identification of borderline situations which could not be categorized.

For example, we have selected the concept “intervenant” (party) to qualify a person who is faced with a conflictual situation concerning his or her residential lease. The “intervenant” could be either the tenant or landlord. The tenant or the landlord belongs respectively to a sub-category such as “sous-locataire” (sub-tenant), “occupant de bonne foi” (occupant in good faith), “conjoint” (spouse), or “concubin” (common-law spouse) on the one hand, or “acquéreur” (purchaser), “mandataire” (mandatary) or “usufruitier” (usufructuary) on the other. Real situations are thus reduced to categories which are normalized in order to correspond to hypothetical conflictual situations. We need to evaluate the impact of this process of reduction and standardization of real

situations. The reduction of legal as well as factual situations, through the categorization of those objects to be formalized, requires a thorough analysis of the acceptable limits of that formalization process.

At the same time, this formalization process reveals that the legal qualification of a real conflictual situation is also a process which reduces it into legal concepts. What must be our main concern is to clearly understand and evaluate the impact of those cumulative reductions of real conflictual situations through the successive categorization processes.

Loge-expert is too small a scale expert system in law to permit such an evaluation at this stage of its development, but we are still justified in our preoccupation with the possible exclusion of borderline conflictual situations, which, however, play a significant part in the interpretation of legal concepts and in their application to diversified real situations.

We have learned from the development of the knowledge base of Loge-expert, that the formalization of legal and non-legal concepts as well as factual ones required to write inferential rules leads, to a reduction of the diversity of real situations into standardized situations. We have named this process the reduction -standardization.

1.4 Stating the implicit

We cannot understand the rules prevailing in the legal field of contracts if the key principles such as mutual consent and freedom of contract are not clearly stated. We mentioned earlier (Figure 1.) that those two principles provide the orientation along which the *Civil Code* articles relevant to obligations are construed, but most of the specific articles related to residential lease constitute exceptions to these key principles, since they are classified as public order clauses which prevail for the negotiation and conclusion of the contract as well as for its execution. Nevertheless, we must refer to those principles to understand the extent to which this contract escapes to their application.

For example, in a non-residential lease, the question of repossession never occurs since

the contract generally ends at the date agreed between the parties at the time of the execution of the lease. In contrast, the residential lease does not end at its term since it is automatically renewed according to the specific principle recognizing the right of tenants to remain in the dwelling as long as they fulfill their obligations. The right to an automatic renewal of a residential lease and the right of the tenant to remain in the premises are principles which contravene those of mutual consent and freedom of contract since parties to a residential lease cannot agree to be released from the latter principles which are mandatory.

To offset those specific principles which have been enacted to protect tenants from unfair evictions, the Civil Code establishes the right of Landlords to the repossession of rented dwellings. The later right is based on the landlords' right of ownership and allows them to put an end to residential leases in order to repossess the dwelling for their personal use or the use of selected members of their family. It is of the utmost importance to set out these opposing legal principles in order to propose a correct formalization of legal concepts relevant to this specific legal expertise.

The formalization of legal knowledge thus requires the explicitation of legal principles and general theories which underlie the meanings of legal concepts.

1.5 The modelization of lawyers' strategies

When a lawyer is faced with a problem of repossession he or she has to analyze the specific components of that kind of problem. The operation may be broken down into successive phases which are linked together along a critical path which protects any from being skipped. We have drawn one possible path which has facilitated the design of interactive communication between the user and the machine as well as the triggering of correct inference rules.

For example, when a landlord wonders about his or her right to repossession of the rented dwelling, the first step is to check if he or she is an eligible landlord to

repossession. We established earlier that all the meanings of the concept “landlord” are not accepted for repossession. Among those meanings, the owner and the usufructuary are eligible landlords whereas the “nu-propriétaire”(‘bare-owner’) is not. The purchaser is a controversial case since legal doctrine and tribunal decisions do not concur in recognizing or refusing his eligibility for repossession. This first step is thus crucial to the success of the landlord’s request .

The second step considers the eligibility of the beneficiary of the repossession. The landlord will be then questioned about the family ties existing with the person to whom he or she will make the dwelling available. Since the Civil Code limits the list of eligible persons to the landlord’s “ascendants” and “descendants”, with an exception for a relative for whom the landlord is the principal support, the eligibility of a beneficiary of the repossession should be thoroughly scrutinized in order to fulfill the Civil Code conditions.

The next step examines whether the landlord owns other dwellings of the same type as the one he or she wants to repossess. If he or she does, the repossession will not be possible and the landlord should take that available dwelling. At that point it is necessary to construe the concept of an “available identical dwelling ”. The Civil Code identifies a dwelling as an identical one if it simultaneously possesses three characteristics: it is of the same type; it has an equivalent rent; it is located in the neighborhood (art.1659.4 C.C.). Tribunal decisions will precise what those criteria mean.

The last step consists of an evaluation of the landlord’s intentions with respect to repossession. His or her good faith will then be weighed according to the proof he or she will bring, since in this issue, good faith is not presumed as is usually the rule in Civil law matters, but must be demonstrated (art.1659.3 C.C.). Tribunal decisions concerning the question of good faith are not very useful to establish criteria of good faith, since some omit the demonstration of good faith even thought the principle of presumed good faith does not apply to a repossession matter. Some decisions do

however find bad faith from former conflictual relationships between the landlord and the tenant.

Step by step, the critical path leads to a complete overview of the most important phases of the repossession process. Some borderline situations have not been covered, such as the case of a landlord who is an undivided co-owner or who intends to transform his rented immovable into an immovable in co-ownership. Those specific situations will be developed in further ad hoc modules, since the technology of expert systems gives us the flexibility to do so.

We succeed in designing a model of the simulated strategy a lawyer could adopt in analyzing a situation related to repossession. Our model is not definitive since more research needs to be done to establish the meanings of legal and non-legal concepts such as “acquéreur” (purchaser), “bonne foi” (good faith), “principal soutien” (principal support), “logement de même type” (dwelling of the same type), “loyer équivalent” (similar rent), “situé dans les environs” (situated in the neighborhood). Nevertheless, it helps us in developing our small-scale expert system and Loge-expert does give correct answers to the user. Figure 3. illustrates this simulated strategy.

[Insert Figure 3 about here]

1.6 The independence of the Loge-expert knowledge base from its interface devices

Since we propose to develop a computerized tool to assist legal decision-making for non-lawyers, the interdependency between the knowledge base and the communication devices with the user is a major issue. We have considered whether the formalization of legal knowledge is related to the characteristics of the end user, but we

could not answer that question until we designed the modules of communication with the user.

On the other hand, we were formalizing the legal knowledge in line with the standard rules usually adopted in the legal profession. Setting out legal concepts was done with reference to legal language prevalent in this specific legal field. We were well aware that non-lawyers are not familiar with this language. Three options were available: to leave the legal concepts non available for a non-lawyer user; to let those concepts available in legal language; to explain them in order to have them understood by non-lawyers. Since legal reasoning is based on legal language, we could not avoid the necessity of having them available to the user, but nor could we sacrifice the constraints of legal language by translating it in plain language. We have adopted a solution which uses the technology of Hypertext to resolve this difficult issue.

Our strategy was, on one hand to isolate the "legal nucleus" which consists of the dictionary of concepts and the inference rules which are triggered according to the simulated research strategy of a lawyer, and on the other hand, to develop communicational layers which convey the explanations of those concepts then expressed in plain language. Those explanations written in plain language aim at making legal concepts understandable by non-lawyer as well as at encouraging learning of legal language in order to get them familiarize them with it, should they consult a lawyer later on with regard to their conflictual situation.

We are now able to conclude that the communicational layers and the Loge-expert knowledge base are independent. We may go on concomitantly with the formalization of legal concepts in legal language in order to improve the Loge-expert knowledge base, as well as with the writing of explanations in plain language in the communication layers to make legal concepts understood by non-lawyers. Consequently, non-lawyers access alternative explanations according to their level of understanding of legal language through the communicational layers.

1.7 The validation of legal concepts through reference to a legal textual data base.

We illustrated earlier (cf. figure 2) the different sources of legal knowledge from which our specific legal expertise will be drawn: statutes and legal provisions, regulations, tribunal decisions, legal doctrine as well as the know-how of practitioners. In order to cope with this ambitious project, it became evident that the development of a legal textual data base would be necessary to gather those heterogeneous documents. How would those documents be analyzed by the computer? We were well aware that the introduction of such textual documents on an automated support is not sufficient to get them ready to be integrated into the knowledge base. We undertook some research about reading models of legal textual documents to understand what the computer could do and what is left to the legal expert (Thomasset, Blanchard, Paquin, 1990).

We were interested by the classification of legal documents established by J. Wroblewsky (1988). This author considers that legal language belongs to natural language and differs from it only because it includes specific legal concepts, which permits its classification among specialized natural languages. Differences between legal and natural languages are semantic rather than syntactic. In his further considerations of legal language, Wroblewsky establishes that legal language should be differentiate from meta-legal discourses. According to that distinction, legal language is expressed in statutes and regulations whereas meta legal discourses are expressed in court and tribunal decisions as well as into legal doctrine and all kinds of discourses about law. It became clear that our legal textual data base includes both categories.

We were then preoccupied with determining if a legal expert reads both categories with the same reading pattern or if he or she adopts specific ones for each category. We recognized two different reading patterns according to the category to which a legal document belongs. We tested these reading patterns with textual documents belonging

to both categories, such as articles of the Civil Code relevant to repossession, as well as tribunal decisions issued in our specific legal field. Each reading pattern could be broken down into successive levels which are specific to each category of reading pattern of legal documents. We listed six levels for each reading pattern such as a syntagmatic level, a paradigmatic level, the correspondence analysis level, the generalization level, the symbolic level. Those levels start from a surface reading to reach a deep level of reading at the sixth level. We have concluded that a computer could be of some help only for the syntagmatic and paradigmatic levels since it can perform a thorough and time-saving search into textual documents at those levels (Thomasset, Blanchard, Paquin, 1990).

We have developed our legal textual data base according to these reading patterns. At the moment, our data base includes only tribunal decisions which belong to the second category established by Wroblewsky. In further development we will add documents belonging to the first category.

Our legal textual data base is also used to help in writing the explanations of legal concepts introduced into the communicational layers. We refer to those textual documents to extract pertinent parts to illustrate some meanings of legal concepts.

2 The general public access to Loge-expert

2.1. The problematic

The goal of making a legal expertise accessible to the general public adds two additional difficulties to the knowledge-based system building. First of all, the user should be motivated to undertake his search for legal expertise without the requirement of any technological knowledge. The general population is not interested in learning how to use a computer to get access to legal expertise. People want to have access to legal expertise through automated tellers as they do at the bank. The problem is that much

more wording is necessary for this type of consultation than to get money or to pay the electricity bill. Secondly, besides the job of legal knowledge modelization and formalization, it must also be translated into plain language. To a certain extent, communication of legal knowledge to the uninitiated could be considered as a special field in itself. In the lawyer's interaction with his or her client, a double translation occurs. First, some facts are extracted from the "story" expressed in plain language by the client and translated into legal language (qualification). Afterwards, the legal reasoning is carried out and the results are translated back into plain language understandable by the layman. In the following pages, we present our thoughts on this communication process and the (partial) solutions we implemented in the Loge-expert prototype.

2.2. The means of communication between the man and the machine

There are three broad means of man-machine communication (Rialle, V., 1988). The first is the command language. We still find it in some computer-operating systems like DOS or UNIX. It offers fixed and ordered expressions, commands and parameters which require a high capacity for abstraction from the user. For example, the command: DIR A: which commands the computer to list all the files of the disk driver identified by the letter A. If this mean of communication with the computer is the easiest from the programmer's point of view, such is not the case for the layman user because few people have the benefit of the necessary level of computer literacy.

The second is natural language statements. For example:

1) GIVE ME THE CONTENT OF THE EXTERNAL DISK DRIVE?

2) I WANT THE LIST OF ALL THE FILES OF THE DISK DRIVE A.

The aim is to have the machine establish a dialogue with the user like a human being. This level of interaction needs more than recognition of key words that is, a deep

semantic comprehension. Let us imagine the complexity of the task of having the computer understand that the wordings of examples 1 and 2 mean the same thing. There are numerous difficulties related to natural language processing, among which are spelling mistakes, approximative synonymy, polysemy, homography, elisions, etc. Several constituents are necessary:

- a phonetizer
- a morphological analyzer
- a syntactical parser
- semantic features dictionaries
- knowledge bases

In spite of all of these constituents, there is still fundamental research going on. But even if all the remaining theoretical difficulties were solved, the benefit of flexibility in the expression would be offset by certain negative points. Until speech recognition is functional, it implies the use of a keyboard which is not necessarily attractive for people untrained in the use of typewriters or word processors. Moreover, the complexity of functions needed to process unrestricted natural language statements is incompatible with the size of the micro computer we are aiming at. Furthermore the processing time will be considered too long; a user does not generally accept a delay of more than a few seconds.

The last category of man-machine communication is graphical. This type of interaction, made possible by the bitmap screen development is based on a general open-system, user-friendly philosophy. This being the case, the dialogue is mainly menu-driven and the user input is generally entered via mouse-clicks (or other type of pointer like finger on a tactile screen) and simple form-filling procedures. It is called graphical because numerous display conventions are offered: multi-window, menu bar, icons, buttons, scroll bar, etc. All the available commands are offered for the user to select by means of the pointer, making the keyboard seldom necessary. Furthermore, it arouses the user's

interest because of the animation: selected, the buttons are inverted, the menu option blinks, etc.

Two main reasons justify our choice of graphical man-machine communication. It offers greater accessibility because there is no prerequisite learning, even though a graphical screen succession might become somewhat tiresome. As the options are all listed in the menu, there are no alternative formulations on misspelling possibilities. By itself this is not sufficient to fulfill our goal of giving people autonomy in consulting Loge-expert. Each option to be selected in the menus or dialogues, each question asked by the system or statements emitted should be translated into plain language.

In the following sections, we will advocate first the separation of the reasoning function of the knowledge-based system from the communicational function with the user and the multiplication of the latter according to the classes of user.

2.3 The separation of the communication function from the legal reasoning

Although a large part of the knowledge-based system runs independently on a micro-computer platform, its input-output (I/O) function, also called man-machine interface, is, in most cases, irremediably mixed with the other functions of the system. While the knowledge-based system is, in some respects, different from the conventional system, we think it could benefit from the evolution of the latter. Thus the I/O function has been transferred from the programmer to the operating system in an independent set of routines (Ralston, A., Meek, L 1976). This migration avoids the accidental destruction of files and writing in the program workspace. Maintainability and evolution of the man-machine interface towards the program itself have become software engineering principles that could improve the knowledge-based system's development.

Since a layman user could not generally understand the legal concepts as they are modelized in the knowledge base without getting some explanation as to their

meanings, special attention should be paid to this function. This is why we extended the above principle to the design of Loge-expert, the expertise (the legal concepts and the legal reasoning based on them) and the interaction with the layman user are kept apart (the mechanism and the textual content). The reasoning in terms of inference rules, the question asked of the user and the legal statement production are made within what we called the "legal nucleus" of the system. By interaction we mean the elicitation of the legal nucleus components in plain language. We called this function a communicational layer. To enable the greatest accessibility from expert to non-expert users, without interfering with the legal reasoning or with the speed of the answers, the independence of the legal nucleus from the communicational layer must be established.

By doing so, the building of the system could be divided between two teams, each having its own speciality: knowledge engineering for the legal nucleus, linguistic and psycho-sociology of communication for the communicational layer. A specific solution could then be coined according to the problems particular to each of those components. Mapping the legal concepts, questions and statements with their plain language elicitation is done simply by means of a table of strings. The communicational layers must be accessible at anytime upon request from the user.

2.4 Why a unique communicational layer?

The uniqueness of the user knowledge-based system brings to mind the execution of a mainframe computer program. The man-machine communication is done by means of commands and parameters, first punched on cards and later written into files with a line editor from the keyboard of a console. Programmers and users communicate with the computer the same way as illustrated below:

[Insert Figure 4 about here]

This is frustrating for users because they have to learn an artificial way to formulate their request. The result of research in the field of information science has showed that the frustration is proportional to the extent of intermediary steps the user must carry out between the expression of a problem and the access to the relevant information Furthermore, if the frustration level is too high, the system, even if it appears to be efficient, will be rejected. The satisfaction survey following the installation of the first large bibliographical databases in the library demonstrate that users are not a homogeneous group. Consequently, a unique man-machine interface does not satisfy all users classes because their needs are somewhat contradictory. The occasional users want to be guided step by step, while the others want to have free access to all the system's resources. Some are very familiar with the system's field, while others not at all, etc. The solution is to respect the users' heterogeneity as much as possible. After a user needs survey, groupings are made and a specific man-machine interface is designed for each of the user categories. Figure 5 illustrates this:

[Insert Figure 5 about here]

A, B, and C represent different man-machine interfaces that could be in the form of hardware (chips) as well as software.

Adapted to our design, this principle enables us to create as needed several communicational layers optimally tailored to the user classes which could be added at any moment during or after building the legal nucleus. The following illustration shows a general communicational layer for the layman user, another for the tenant association user who has a good background in the legal concepts which concern him or her and

none at all for the relative to the opposite position.

[Insert Figure 6 about here]

Nevertheless it might be difficult to determine which legal concepts are controlled by a specific class of user. The grouping of users on the basis of the similarity and divergence of the need expressed may be difficult. Both operations must rest on a rigorous inquiry based on validated questionnaires administered to a representative sample.

2.5 The ergonomics of the communicational layers

The aim of Loge-expert is to motivate a legal and computer uninitiate to solve his legal information queries by himself. The communicational layers are designed to facilitate the interface between a layman and the legal reasoning. Indeed, the layman does not have the same motivation as the experts; he is usually not ready to undertake learning a complex system. In order to succeed in matching human being, legal knowledge and computer, the socio-cultural characteristics of the users have to be taken into account in the ergonomics of building the communicational layers.

Knowledge-based system ergonomics are more developed in industry than in the legal domain (De Terssac, G., 1988, Barthet, M.F., 1987). Very little has been done to understand the way laymen are faced with computers to get legal knowledge. Much more research has to be directed towards adapting technology to human constraints. Instead of training people to adapt themselves to computer languages, we should find easier ways of helping people interact with knowledge modelized into computers. This research requires a multidisciplinary approach which includes psychology, information sciences, communication sciences, socio-linguistics, computer engineering as well as

marketing. Furthermore, the success of such research depends partially on the social commitment of the research team. Among the computer science research sectors, the one oriented towards the final user seems to be the most promising for our purposes: the adaptation of the technology to its user.

Knowledge-based systems have improved much more on the side of software engineering than on the user's side. We know very little about the general public as users. We need to understand their behavior when they are seeking information. We should pay attention to multiple human factors such as the user's profile, his cultural background, his values, his level of expertise knowledge. Prerequisites for the use of those systems, such as training to run them or to understand the wording of a given answer, should be finely tuned to the expectations of the user, since this could be as important a factor in the system's success as the consistency of the legal nucleus.

Data on users should be contextualized. For example, observations of lawyer-client interviews should be made in order to understand the way questions are asked and answers are given by both parties. Even if asking questions is the most common day-to-day human activity, the technical understanding of it is limited (Kerpley, G.P., 1976). Our observations will thus help us to understand this process in the very specific situation of a legal field. The mechanism which structures the formulation of questions should be fully understood because it provides an articulation for the legal knowledge modelization. Furthermore, special attention should be paid to strategies which, during the legal problem resolution process, discriminate among clients' answers to avoid bias and prejudiced information.

2.6 The wording of the communicational layers

The ergonomics of a given system should not be only a technological issue but also a linguistic one because even if the system is made easy to work, if the wording of its objects, actions and instructions are not easily understandable, it could not be

considered user friendly. Thus the complexity and the diversity of the wording of the concepts, questions and statements involved in the legal nucleus of Loge-expert should be reduced to a minimum. The vocabulary and the syntax should have been previously normalized. The wording should be designed with state of the art techniques developed on "plain language" and readability of legal language (Ribordy, F.X., Laflamme, S., Cazebon, B., 1986-1987).

After having set up the chaining of the questions and explored the admissibility of the system's answers from a layman's point of view, we should focus on the language used for the elicitation of the legal reasoning. Two aspects have to be covered: the lexicon and the wordings. For each legal concept, we will find one or more equivalent expressions understandable by laymen. This lexicon will have to be validated by both lawyers and laymen. The wordings of the definition of legal concepts should take into account the more recent research on readability (Fernbach,1990; The Plain Language Centre 1990). From a practical point of view, we will elaborate a set of rules on the length of words, the complexity of sentences in terms of propositions etc.., to design the wordings.

This methodology will prevent us from arriving at ad hoc solutions and will enable us to distribute the task of designing the wordings among several members of the research team on the one hand, and to spread it over a long span of time on the other. Once the wordings are drafted, we will turn to their visual representation on the computer screen, because a screen window is a completely different information media from paper. We will take into account visual semiology analyses of the structure and the components of the messages and its representation. As we have opted for a graphical man-machine interface, we should be attentive to the scientific and technological developments in these aspects of computation.

2.7 The implementation of the communicational layers

Instead of implementing the communicational layers design in a deterministic way, that is, according to the user category as determined at the beginning of the consultation, with the corresponding explanations given to the user, we have chosen a more flexible formula. The explanation about what is going on in the legal nucleus should be given to the user anytime but only upon his request and only up to the level he desires. We have chosen the hypertext technology to implement the communicational layers because it offers those possibilities. We developed a function, called "hyper_aide", available anytime during the consultation, by displaying a card presenting relevant textual or graphical information in a specific window. This function is activated by a hot key or pointing to a specific icon and its aim is to answer such usual requests as: "Why is the system asking me this question?" "How does the system draw this conclusion?" We add the elicitation request "What does this concept mean?". The cards are structured according to a model of concentric circles:

[Insert Figure 7 about here]

First, a short explanation is given, then, upon request, more information or an extensive explanation could be accessed:

[Insert Figure 8 about here]

Furthermore, the cards are linked by means of a menu presenting the related concepts

as shown in the previous illustration. These links enable us to more easily delimit the card content boundaries without stopping the information associations.

For example, a card explaining a concept, here "locateur", will include the following fields: a) a plain language translation of the concept and a succinct definition ("Définition"); b) some references to complete the previous definition ("Références") and c) the decisions illustrating this concept ("Jurisprudence"); d) the links with the related concepts ("Voir aussi").

[Insert Figure 9 about here]

[Insert Figure 10 about here]

But the hyper_aide function covers only one part of our needs: the explanation in the user's language of the consultation questions and results which are expressed in legal language. The printing of a brief report should be added. This report would summarize the legal consultation and the correlative explanation. It could be a useful tool for the user to remember the consultation, to compare with the results of other inquiries, to share the results with others, or to facilitate mediation.

2.8.Closing remarks

We are aware that the value of the plain language explanation of legal concepts will remain relative, even if the cards are written with care. Comprehension depends on the user's ability to integrate them in his or her cognitive structure. This is why the

hyper_aide cards are structured in several fields and one of them should show the conceptual structure of the field, the context. This aspect, in our view, is as important as definitions. The hyper_aide function could be complemented by request to a full text database whose content, in this case, would be the relevant Civil code articles, a representative amount of decisions and other relevant texts (doctrine).

We thought that it is better not to conceal the legal content of the system, and instead to elicit it upon the user's request. In this way our system has a pedagogical side effect: it enables a layperson to be part of the solution to his or her legal problems.

3 Textual data base

At first conceived as a database made of case law, the textual database has rapidly became a multipurpose documentary repository. The first task that it has to fulfill is to make all the decisions from the "Regie du Logement" in the selected area available. This is necessary because no on-line servers allow us access to these cases. The second and complementary function has to do with validation of the rules in the knowledge base from real cases in the same area. The third function, as we have previously seen, provides explanations to facilitate the understanding of legal concepts triggered by a user's consultation of Loge-expert.

Now we know that this textual database will consist not only of cases but of other texts necessary for the knowledge base or for the communicational layer, the database will be expanded as necessary.

3.1 The aim: documentation

In order to provide documentation, to support inference rules and also to validate and illustrate them, we had planned to gather and to analyze all the decisions of the Regie

du Logement, the tribunal of first instance that generates the cases which are of most interest to us. Unfortunately, the Regie du logement only publishes a small number of its decisions, about one hundred and fifty annually, which is less then a hundredth of the cases heard in a given year.

The more traditional legal data banks, that are available on-line through Soquij or Quicklaw (+ Westlaw for Canada...) does not meet our requirements: they are based on published material. Thus, only the decisions of the Regie du Logement that are appealed before the Quebec' court or are evocated before the Superior Court will, perhaps, show up in these on-line databases. On the other hand, these on-line servers are available only if you can master the communication process which is slow and / or costly. Realistically, they cannot be consulted in parallel with an expert system.

In fact, they can only be made available after the user has learned how to telecommunicate, which requires a functional knowledge of a terminal, and mastery of the inquiry process itself, based on a specific query language. Those query languages are proper to each on-line server and each of them begs for a certain familiarity before being put to its full capacity. We cannot expect this level of expertise or patience from laymen users. Moreover, these servers bill the users mainly on a connect-time basis, representing an uncontrolled cost, over and above the cost of exploitation of an expert-system.

All the above, make their consultations unsuitable for building an expert system.

3.2 Building the data base.

Our aim of building a database of case law was simple only at first sight. The decisions from the Regie du logement are not available in a computer-readable format, few of them are published and those that are, only after an internal screening process based on criteria that are unofficial.

In fact, we could describe the materials for our database in terms of the conjunction of three strata:

—From 1983 to 1986, the Regie du Logement did not publish any of its decisions in full but merely summaries of some selected ones. The other decisions were archived under the civic address of the premises that was litigious; those decisions are then less accessible.

—From 1987, the Regie du Logement does publish the full text of those decisions selected in a screening process which involves such criteria as the extent of the reasons, novelty or the consolidation of a jurisprudential trend.

—Then, beginning in the autumn of 1990, we received all the original texts of decisions relating to repossession and rent reduction, under an agreement with the Regie du Logement.

We decided to translate some of those decisions into machine readable format by means of an optical character recognition system. Our goal is still to establish a database of all the relevant decisions even if we have to restrict ourselves to the most readable ones: that is to say the decisions that are not spoiled by the copying process. Then after this "reading" the decisions still have to be verified, analyzed and saved in a database.

Those operations are possible in part because of the use of some exploring techniques, which we will not describe here in detail. Let us simply remind you for instance that any search in a textual database is not as simple as in a structured and normalized database: texts do not fit well in limited fields and the thorough analysis of any legal decision is a task which surpasses our immediate capacity.

Instead, we opted for a simpler treatment of the texts based on certain more obvious characteristics such as the reference number or the type of the party:

[Insert Figure 11 about here]

We must keep in mind that unpublished decisions are not easily categorized: even their identification is more complex than usual [they lack unambiguous reference numbers and of course bibliographical information like page numbers...] so that the process of extracting key-words requires an intelligent reading.

3.3 Tools to explore the database

We then opted for tools that could be applied to texts after only a minimalist treatment in order to ascertain their completeness, but which nonetheless preserve their associative quality.

Decisions are lumped into one file and then once indexed, a browser acts like a "key word in context" apparatus:

[Insert Figure 12 about here]

Here, the software tool is "Free Text" from Mark Zimmerman. It permits us to localize any given word within its limited and its expanded immediate context. The user interface is divided into three zones:

—the upper part consists of an alphabetical list of words, defined as characters between blanks, with the number of their occurrences,

—the limited context of seven words,

—and in the lower part, the extended context where one selected word is set against its background.

It should be noted that the three zones are not limited: you can move in it so as to point to any word.

Here, we have selected the word "usufruit" [usufructuary] as an example of the consultation process.

[Insert Figure 13 about here]

By the same process we can also find all the occurrences of other related terms like "usufruitier" or "usufruitiere" and then look at the contexts of their utilization.

[Insert Figure 14 about here]

It is then possible to search for and locate some of the rules that are formulated by the adjudicators in the process; rules like:

"le tribunal se rend aux arguments de l'avocate et conclut que seule l'usufruitière... a droit à la reprise de possession du logement"

A rule that could be translated as: **Only the usufructuary, whenever one exists, has the right to repossession.**

It is also possible to feed decisions into a co-occurrence analysis engine that can draw maps, based on frequency of co-occurrences, of related concepts.:

[Insert Figure 15 about here]

This map is obviously related to the concept of ownership. These maps, drawn with Candide™[**], the Hume's machine, are but one example of a quasi-automatic reading of texts. They are related to the extent of the vocabulary that is fed into the lexical indexer of Candide™, but they are also bound by the weight of texts, ie. by the number of decisions that are being indexed.. They can be manipulated, centered around an important concept for instance, or they can~ when they are drawn in a historical sequence, help to focus on a new interpretation, a new trend.

3.4 The feedback to the knowledge-base.

The decisions database can serve to explain, to validate the expertise, but it could also serve to expand it. New cases, usual or unusual, provide new facts with perhaps new applications of rules or even new sub-rules.

3.5 The feedback toward the end-user.

The decisions database also serves to illustrate some of the situations described by the end-user whenever he or she consults Loge-expert by means of cases. Retaining the values that are chosen by the user, it is then possible to locate some of the decisions that embodied those same values so that a comparison may be made.

Prior to that, we also retain some of the wording used in decisions in order to clarify our explanatory messages in the communicational layers.

[**] Candide™:
- Centre de Sociologie de l'Innovation, 64 bvd St Michel, 75006 PARIS; (CSI)
- Institut de l'Information Scientifique et Technique, 26 rue Boyer 75020 PARIS. (InIST)

Conclusion

From this two-year experiment with Loge-expert, we have learned that the development of an expert system in law is an iterative operation which includes simultaneous exploration of different dimensions such as the conceptualization of the specific legal field as well as the conceptualization of the communication devices between that legal expertise and the user or the conceptualization of the linguistic and textual dimensions of that legal expertise. At the moment we have limited our exploration to one generator of expert systems, D_Expert. We know now that whatever our technological choice may be, we must start with the modelization of the legal knowledge in order to be able to formalize it into a specific technological support. That preliminary stage depends to some extent on the technological choice, but the formalization of the legal knowledge can be exported to other types of expert systems shells, and the crucial issues are not related to technological choices.

The modelization of legal knowledge raises some crucial issues since the attempt to formalize it into cognitive structures proposed by a generator of expert systems reveals problems which present legal dimensions rather than technological ones. For example, it becomes clear, that law offers open textures as well as general principles with extensive interpretations; that court and tribunal decisions which were considered as establishing the interpretation of some legal concepts are replaced by more recent ones; that legal language is a barrier to the clear understanding of legal rules. Technological choices could not be a substitute for those issues specific to law, and no shell or programming languages could propose magic solutions to solve that kind of difficulty.

On the other hand, we know that the identification of these issues, could be of some help in order to elucidate their exact dimensions as well as to propose research to

evaluate whether the solutions are possible. At that point, lawyers who are faced with the incapacity of technological choices to solve the difficulties they are used to facing in their day-to-day practice should pay attention to a necessary thorough analysis of their raison d'être.

The development of Loge-expert has led us to the conclusion that most difficulties related to the specific nature of law can be resolved with technological solutions. For example, we could break down the legal expertise into successive modules in order to formalize all the legal concepts required for the treatment of specific situations. Statistical calculus offers solutions to the uncertainty of court and tribunal decisions since we can thereby establish tendencies which can be graded. Legal language could become accessible through explanatory notes expressed in plain language and integrated into hypertextual devices allowing non-lawyer users access to legal documents. Our experiment reveals that every time we were faced with a problem in the development of Loge-expert, we finally found a technological solution since we had thoroughly analyzed its legal dimensions.

Those considerations lead us to conclude that lawyers have an important part to play in the development process of an expert system in law. The computerization of legal knowledge includes a deep reflection about law and the related practices. Law being faced to A.I. technologies, should be thoroughly scrutinized in order to evaluate the opportunity to integrate new technological solutions into law practice as well as the limits of the specificity of its nature.

References:

BARTHET, M.F. (1987). *Ergonomie du logiciel*. Paris, France: Dunod.

FERNBACH, NICOLE, (1990). *La lisibilité dans la rédaction juridique au Québec.* Le Centre de promotion de la lisibilité. Toronto, Canada: CLIC.

PAQUIN, L.-C. (1987). DÉRÉDEC-EXPERT, outil d'ingénierie cognitive. *International Congress on Terminilogy and Knowledge Engineering*. Trier, FRG.

PAQUIN, L.-C. (1987). DÉRÉDEC-EXPERT, environnement informatique pour la génération de systèmes experts. Présentation primée au *Colloque sur les systèmes experts et leurs applications*, Avignon, Fance.

THE PLAIN LANGUAGE CENTRE & CANADIAN LAW INFORMATION COUNCIL, (eds.) (1990). *Plain Language Ressource Materials*. Toronto, Canada: CLIC.

RALSTON, A., & MEEK, L. (1976). *Encyclopedia of Computer Science*. New York, N.Y.: Van Nostrand Reinhold Company.

RIALLE, V. (1988). Ergonomie cognitive et bases de connaissances. *Le Journal de l'Intelligence Artificielle*, (9), 4-9.

RIBORDY, F. , LAFLAMME, S., & CAZEBON, B. (1986-1987). Les textes de lois sont-ils lisibles et compréhensibles? *Revue de l'Institut de sociologie*, 2(1), 223-244.

TERSSAC, G. de, SOUBIE, J.-L., & NEVEU, J.-P. (1988). Technologie et organisation: les savoirs menacés? *Sociologie du Travail*. (3), 462-477.

THOMASSET, C. (1988) Expert System and Legal Formalization: Evaluation of a Prototype in Québec Housing Law. *The International Conference on Intelligence and Society*. Vienna, Austria: International Social Science Council, European Coordination Centre for Research and Documentation in Social Sciences.

THOMASSET, C. (1989). Expert System in Québec Housing Law: from HOME-Expert I to HOME-Expert II. *The International Conference on Law and Artificial Intelligence: Expert Systems in Law*, Bologna, Italia: CIRFID.

THOMASSET, C., BLANCHARD, F., & HEBERT, R. (1988). L'informatisation du savoir juridique: conception d'un prototype de système expert en droit du logement. in Caron, A., & Michaud,A., (eds.), *Recueil des activités du CIEST 1987-1988*, (301-314). Montréal, Canada: UQAM.

THOMASSET, C., BLANCHARD, F., & PAQUIN, L.-C. (1990). Legal knowledge elicitation: from

textual databases to expert systems. *DEXA '90, Data Base and Expert Systems Applications,* (167-173). Wien, New-York: Springler-Verlag.

THOMASSET, C., HÉBERT, R. (1988). La démocratisation du savoir juridique par son informatisation, virtualités, contraintes, effets. in Laperrière, R., Claret, H., & Péladeau, P., (eds), *Une démocratie technologique?* (343-358). Montréal, Canada: ACFAS-GRID.

THOMASSET,C., & PAQUIN, L.-C.(1989). Expert Systems in Law and the Representation of Legal Knowledge: Can we Isolate it from the Why and the Who?, *III International Congress: Logica, Informatica, Diritto: Expert Systems in Law*. (751-772), Firenze, Italia: Consiglio Nazionale delle Ricerche, Istituto per la documentazione giuridica.

WROBLEWSKY, J. (1988). Les langages juridiques: une typologie. *Droit et Société, Revue internationale de théorie du droit et de sociologie juridique*. (8),13-27.