

# Une lecture experte des textes juridiques

**LOUIS-CLAUDE PAQUIN**

Centre ATO•CI

Université du Québec à Montréal

Case postale 8888, succursale A

Montréal (Québec)

CANADA H3C 3P8

Tel: (514) 987-6189

Fax: (514) 987-4567

E-mail: PAQUIN@ATOCL.UQAM.CA

## 1. Introduction

Les textes constituent à la fois le matériau premier et la finalité des juristes. Un avis légal est émis suite à un raisonnement juridique mené à partir d'informations provenant de textes juridiques. Le volume de l'ensemble de ces textes juridiques est toujours imposant. Au Québec, l'ensemble des lois refondues (1985) comporte plus de 520 lois contenant au-delà de 80 000 articles; à ces lois s'ajoutent au moins 2 000 règlements. On estime, de plus à 50 000 le nombre de jugements rendus par les différents tribunaux judiciaires du Québec.

Dans la foulée des développements de l'informatique documentaire à la fin des années soixante, les sciences juridiques ont décelé le potentiel de l'ordinateur pour gérer et accéder aux textes pertinents à l'exercice de leur expertise. C'est ainsi que la Société québécoise d'information juridique (SOQUIJ) a été mise sur pied en 1976 avec le mandat de «promouvoir la recherche, le traitement et le développement de l'information juridique en vue d'en améliorer l'accessibilité au profit de la collectivité» [Montpetit, 1992:1]. Le très grand volume de textes, de même que la technologie disponible posent des problèmes importants quant à l'adéquation des segments de textes repérés par rapport à la requête d'information qui est formulée par le juriste.

Par ailleurs, le regain d'intérêt pour l'intelligence artificielle dans les années soixante-dix qui a, entre autres, donné les systèmes experts et les modèles de représentation des connaissances a eu des répercussions sur les sciences juridiques. De nombreux projets de construction de systèmes experts appliqués à des domaines du droit ont été entrepris, plusieurs congrès d'envergure ont eu lieu. Les publications qui ont

accompagné ou suivi ces projets ont mis à jour un certain nombre de problèmes relatifs aux systèmes experts en tant que dispositifs informatiques et à leur finalité qui était la modélisation des processus décisionnels particuliers aux différents domaines du droit.

Dans la présente contribution<sup>1</sup>, nous allons d'abord examiner la problématique rattachée à ces deux technologies informatiques pour en dégager une fonction qui, dans les deux cas, s'avère nodale : l'interprétation ou l'accès au contenu conceptuel des textes juridiques. Puis, nous allons rappeler le modèle computationnel de la lecture experte qui permet d'assister cette fonction interprétative. Enfin, nous allons tenter d'illustrer notre propos par la modélisation de l'interprétation d'un concept flou : la bonne foi dans le contexte de la reprise de possession d'un logement par le locateur.

## 2. Problématique reliée au recours à l'ordinateur

### 2.1 Les bases de données textuelles

La mise au point de périphériques permettant un stockage de grandes masses de données à faible coût a permis l'extension aux textes intégraux de la technologie des bases de données, utilisée dans un premier temps pour les références bibliographiques. Pour une historique de la recherche textuelle dans le

---

<sup>1</sup> Ce texte est sous presse dans un volume édité par Claude Thomasset dont le titre est: *Les Sciences du texte juridique: le droit saisi par l'ordinateur*, Éditions Yvon Blais.

domaine du droit, voir [Bing, 1984]. La technologie des bases de données textuelles couplée à celle des réseaux de télécommunication rendent possible un repérage en ligne (*on line*) d'information dans les textes juridiques. Les bénéfices escomptés étaient la réduction du volume occupé par les textes juridiques, l'assurance que la version consultée est la plus récente, une accélération du temps requis pour leur consultation, de même qu'une accessibilité accrue. Il devenait possible à plusieurs personnes de consulter en même temps dans des lieux différents les mêmes documents juridiques. Toutefois, ces bénéfices n'ont pas encore surpassé les inconvénients ou irritants causés par le recours à l'ordinateur pour les juristes. Ces irritants sont moins liés à la spécificité du texte juridique qu'à la technologie elle-même.

Jusqu'à tout récemment, pour effectuer un repérage en ligne d'informations pertinentes, le juriste devait formuler sa requête dans les termes d'un langage d'interrogation qui était prévu pour optimiser la recherche au détriment de l'aisance d'utilisation. L'utilisation d'opérateurs booléens pour établir une stratégie de recherche efficace nécessite une familiarisation sinon un entraînement de la part de l'utilisateur. Plusieurs efforts ont été faits pour pallier à ce problème; parmi les solutions explorées, mentionnons la formulation des requêtes en langue naturelle et le dialogue dirigé par un système expert qui assiste l'utilisateur dans la formulation de sa requête [Côté, 1992].

Deux mesures permettent d'évaluer l'efficacité du repérage de l'information étant donnée une requête : les taux de rappel et de précision. Le taux de rappel mesure la proportion de documents pertinents qui n'ont pas été repérés; alors que le taux de précision indique la proportion de documents non-pertinents qui ont été repérés. Le mode de repérage dit «plein texte» présente les pires taux. Seuls les passages de texte présentant les mots de la requête sont repérés, de sorte que lorsque ceux-ci sont substitués par un synonyme ou un pronom, les passages sont ignorés. De plus, rarement toutes les occurrences des mots sont pertinentes, soit parce que ceux-ci présentent plusieurs acceptions ou parce que les thématiques diffèrent.

Pour améliorer la performance du repérage, le recours à une forme ou une autre d'indexation s'avère essentielle. L'indexation, dont il est ici question, n'est pas l'assignation d'une adresse aux mots du texte afin d'accélérer le repérage mais bien une catégorisation de segments du texte au moyen de descripteurs qui sont structurés en un thésaurus. La nécessité d'un accès conceptuel et analogique à l'information contenue dans les textes juridiques a été démontrée par Ashley qui a proposé une méthode d'indexation par dimensions ou caractéristiques déterminées importantes par la cour pour fonder la décision [Ashley, 1988]. Comme cette contribution ne porte pas comme telle sur l'indexation des textes juridiques, nous nous contentons de mentionner qu'il s'agit d'une opération cognitivement complexe qui dans tous les cas, met en jeu une dimension d'analyse et surtout d'interprétation du contenu des textes. S. Bertrand-Gastaldy [1990] fournit une présentation détaillée de la problématique l'indexation. Il va sans dire que cette opération, qui nécessite un personnel hautement qualifié, engendre des coûts élevés, de sorte qu'il faut un marché potentiel suffisant pour justifier cette entreprise.

La pratique de l'indexation est bien antérieure à l'avènement des bases de données textuelles. En effet, ce que l'on pourrait appeler la «chaîne de traitement jurisprudentielle» comporte, dans l'ordre, les opérations suivantes : la sélection, la classification, le résumé et l'indexation des décisions motivées jugées pertinentes. Les recueils de jurisprudence comportent un index qui présente en ordre alphabétique, la plupart du temps sur plusieurs niveaux de généralité, la liste des descripteurs utilisés accompagnés de la référence des textes auxquels ils ont été assignés.

Quoiqu'essentielle, puisqu'elle permet un accès aux textes par les concepts peu importe leur expression, cette opération pose deux ordres de problèmes. Le volume des textes qui peuvent être indexés dans un temps donné, dans un cadre où les ressources sont limitées, est de beaucoup inférieur au volume des textes qui pourraient être avec profit diffusés. Dans le contexte québécois, cette question se posera avec acuité lors de la réalisation, par le

ministère de la Justice du Québec, du système de rédaction et de diffusion électronique des jugements [Laferrière, 1992]. Des recherches sont présentement en cours pour développer une assistance informatique à partir d'un modèle de la façon de faire actuelle des indexeurs; le système prendra la forme d'un système expert [Bertrand-Gastaldy, 1992].

Par ailleurs le thésaurus qui est en fait le répertoire structuré des descripteurs qui seront utilisés pour l'indexation et, subséquemment, pour le repérage détermine l'accès aux textes. Il s'agit d'une limitation *a priori* des lectures et relations qui sont le fait des lecteurs; tel est le prix à payer pour une plus grande efficacité. Si un recours aux descripteurs s'avère satisfaisant dans la majorité des cas d'usage courant des textes juridiques, il est des cas où se dessine le besoin d'un dispositif souple et même dynamique qui permette d'étudier une problématique donnée, reliée à l'interprétation d'un concept ou encore d'une règle appartenant au domaine juridique.

## 2.2. LES SYSTÈMES EXPERTS

Dans la foulée des recherches en intelligence artificielle, de nombreuses recherches ont été entreprises en sciences juridiques, notamment dans les domaines de la représentation des connaissances et des systèmes experts. Pour un historique, voir entre autres [Tyree, 1989]. Ces recherches avaient un but plus ambitieux que celui de faciliter l'accès à l'information juridique : simuler le raisonnement (*emulate the reasoning process of a lawyer* [Susskind, 1986: 168]) d'un juriste qui lui permet soit de proposer des solutions à des litiges ou encore de rendre des décisions.

### 2.2.1 La modélisation des connaissances

Cette simulation du raisonnement est effectuée par le biais de règles de production ou de clauses logiques permettant un accès associatif à des faits qui invoqueront d'autres règles ou clauses, etc. La formulation des règles ou des clauses nécessite une phase préalable de modélisation des connaissances. Cette opération s'est avérée cruciale pour la construction des systèmes experts:

Knowledge representation has been identified as the most important issue facing the field of AI and Law [McCarty, 1990: 30]

Dans le domaine juridique, cette connaissance dont on veut faire le modèle recherché réside, pour une bonne part, dans la capacité d'interpréter les textes juridiques. La plupart du temps, le problème de la modélisation de l'interprétation juridique a été envisagé de la façon suivante par les constructeurs de systèmes experts [Kowalski, 1979] :

- 1) en normalisant le contenu même des textes de lois sous la forme de SI prémisse ALORS conclusion;
- 2) en ajoutant une structure de contrôle qui permette la résolution de litiges ou la prise de décision à partir des règles;
- 3) et en ajoutant de l'information heuristique.

La structure de contrôle est en fait le principe de «réduction des problèmes» [Hofstadter, 1979]. Le problème initial est divisé en sous-problèmes de plus en plus élémentaires. Cette division se poursuit jusqu'à ce qu'un niveau soit atteint où les sous-problèmes peuvent être résolus. En PROLOG, la division en sous-problèmes se fait par permutation d'arguments lorsque, dans deux clauses, les prédicats sont les mêmes; cette procédure est appelée «unification». Les sous-problèmes sont résolus par unifications avec des clauses sans prémisse et dont les arguments des prédicats sont des constantes; ce type particulier de clause constitue les faits. Il est apparu très tôt que cette subdivision récursive des problèmes entraînait une explosion combinatoire dont les conséquences allaient de la dégradation de la performance du système jusqu'à l'impossibilité de computation selon la taille de «l'espace du problème» et l'ampleur des ressources computationnelles disponibles. Plusieurs stratégies computationnelles locales ont été développées pour pallier à ce problème, entre autres l'ordonnancement des clauses concurrentes, formant un ensemble de conflit, de façon à ce que celle dont la probabilité de succès est la plus élevée soit déclenchée en premier [McEnery, 1989]. Toutefois, c'est du

côté de la compréhension de la cognition humaine que la solution semble résider. Les humains, au fil de leur pratique, développent des connaissances qui, bien qu'incertaines, leur permettent de raccourcir l'espace de recherche d'une solution à un problème; ces connaissances sont appelées « heuristiques ».

La normalisation des textes de lois sous la forme de règles de production ou de clauses a d'abord été conçue comme le passage (*mapping*) d'une structure de symbole, la langue juridique, à une autre structure, celle de la logique des prédicats du premier ordre en l'occurrence. Une illusion de transparence est créée par l'implication du processus d'interprétation. Or, il s'est très tôt avéré que les objets textuels n'étaient pas directement ni nécessairement réductibles à une logique connue ainsi qu'à des composantes bien délimitées :

In many legal domains it is therefore a mistake to attempt to represent legal knowledge in the surface level language [Kowalski, 1991: 22]

Plusieurs chercheurs, inspirés sans doute des travaux en linguistique, ont démontré l'importance de mettre à jour la structure profonde des textes juridiques. Cette structure profonde, indépendante de la surface des textes, serait, dans le cas de la jurisprudence, constituée de patrons de faits qui justifient les décisions [Smith, 1987]. En plus des problèmes « syntaxiques », la normalisation des textes juridiques pose des problèmes « sémantiques » :

Besides the 'syntactic difficulties' inherent in normalizing statutes in the form of rules, there are 'semantic' difficulties such as the presence of conflicting rules, imprecise terms and incompleteness. [Vossos, 1991:32]

Les règles de droit contradictoires, de même que les concepts flous, sont des règles et des concepts qui, avant d'être transposés, doivent être l'objet d'une interprétation juridique.

Les règles de droit ne sont habituellement pas contradictoires en elles-mêmes, c'est plutôt qu'elles donnent lieu à des interprétations alternatives et que le constructeur du système

expert est à la recherche de l'interprétation unique qui est vraie :

The usual way of constructing legal expert systems is to embody a single structural interpretation of the set of legal rules that are the basis for the system. [Allen and Saxon, 1991: 61]

La plupart des causes de ces interprétations alternatives sont attribuables à des facteurs textuels d'ambiguïté [Allen and Saxon, 1991] dont nous faisons ici l'économie de la nomenclature. La plupart de ces ambiguïtés ne peuvent être levées avec le recours à une analyse linguistique seule. En effet, le texte est le produit d'un faisceau de forces parfois contradictoires, telles :

- la thématization qui met en valeur ce de quoi l'on parle;
- l'anaphorisation qui permet de substituer un terme par un pronom ou un autre terme qui le recatégorise;
- l'intertextualité qui établit des rapports avec d'autres textes apparentés;
- l'interdiscursivité qui met l'accent sur les principaux discours actifs dans un milieu proche de la personne qui écrit.

Dans le cas qui nous occupe, le texte ne sera compris qu'à la suite d'une interprétation juridique.

Par ailleurs, une des composantes de la représentation de la connaissance est la description compositionnelle des concepts en primitives sémantiques. De telles théories logico-algébriques permettent un calcul sur la représentation produite qui peut être soit en schèmes structurés, tels les *frames*, les graphes conceptuels et les scénarios, ou en schèmes construits selon la logique des prédicats du premier ordre. Les concepts flous (*open texture concepts*) [Gardner, 1987] sont les concepts à l'œuvre dans les textes juridiques dont la description ne peut être faite une fois pour toute. Il en est ainsi parce que ces concepts sont dynamiques, et que leur signification change lors de leur application d'un cas à un autre [McCarty, 1981], ou parce que les définitions qui fourniraient les

conditions nécessaires et suffisantes pour leur délimitation sont absentes [McCarty, 1989] ou encore parce qu'il s'agit de concepts trop abstraits [Branting, 1991].

Ces deux phénomènes des règles contradictoires et des concepts flous rencontrés lors de l'ingénierie cognitive mettent en relief la faiblesse d'une sémantique immanente qui considère que les termes du texte réfèrent directement à des concepts et que les énoncés réfèrent directement à des règles juridiques. Une telle sémantique néglige le contexte d'occurrence des termes et des énoncés en les tenant pour de simples variables, alors que pour une sémantique interprétative [Rastier 1989], le contexte participe, dans tous les cas, à la détermination sinon à l'élaboration du sens d'un concept ou d'une règle. Qui plus est, l'équivocité est une donnée fondamentale des textes:

En d'autres termes, et bien que toute notre tradition herméneutique milite contre cette conclusion, le sens d'un texte n'est pas de l'ordre du vrai, mais du plausible. Plutôt que de révoquer les interprétations jugées impropres, il convient donc de la hiérarchiser, en graduant leur plausibilité relativement à une stratégie donnée [Rastier, 1991: 160]

Ainsi, la représentation de la connaissance à partir des textes, juridiques ou autres, demande de déplacer les efforts, généralement tournés vers l'élimination des ambiguïtés, vers l'intégration d'interprétations plurielles. Pour mener à bien une logicisation des énoncés textuels, on devra d'abord colliger et classer les contextes d'occurrence des termes désignant un concept ou énonçant une règle à partir d'un corpus de textes pertinents: lois, jurisprudence et doctrine. Ce classement est une opération cognitive qui nécessite la médiation de la compréhension par un juriste qui, selon les cas, produira une ou plusieurs interprétations. Il y a donc une nette distinction entre les connaissances véhiculées par un texte et les connaissances que requiert son interprétation.

### **2.2.2 Les systèmes experts ne permettent pas un traitement sémantique mais seulement syntaxique de la connaissance**

Les systèmes experts ont été pensés et développés dans le cadre de recherches dans le domaine de l'intelligence artificielle. Leur finalité était anthropomorphique c'est-à-dire imiter le mieux possible le raisonnement humain, au niveau de la performance du moins [Buchanan 1984]. C'est ainsi qu'une confusion entre le formalisme et le substrat auquel il s'applique s'est instaurée et, par la suite, a été cultivée jusqu'à générer des attentes qui, comme nous le verrons plus loin, ne pourront jamais être comblées par un automate. Étant donné que MYCIN, le plus célèbre des systèmes experts, visait le raisonnement clinique dans le domaine des maladies infectieuses du sang et de la méningite, on dira que les systèmes experts font du «diagnostic». L'exécution du système expert s'appellera une «consultation» et la trace d'exécution sera l'«explication». Est-il possible que cette confusion touche les systèmes experts en droit où la logique, comme mode d'accès associatif à l'information, a été confondue avec la logique comme justification du raisonnement juridique; la règle de droit, avec la règle d'inférences?

Une telle confusion explique qu'on a tenté de pallier aux insuccès des systèmes experts à imiter le raisonnement d'un juriste en proposant des remaniements au niveau de la logique du moteur d'inférences ou de la représentation des connaissances. Ces remaniements vont dans le sens d'une logique déontique [Jones, 1990], d'une logique temporelle [Poulin, 1989] ou encore d'une logique modale [Fattibene, 1989], etc. Bien que toutes ces contributions améliorent la qualité ou l'envergure du traitement des symboles, structurés ou non, qui représentent les concepts juridiques, le traitement demeurera toujours syntaxique.

En effet, la logique propositionnelle de même que la logique des prédicats du premier ordre ne permettent que des manipulations

syntaxiques d'où toute signification est totalement exclue. Seul un observateur extérieur est capable, en interprétant une configuration de symboles, de lui conférer une dimension sémantique. Pour Dennet, la circularité ou l'enchevêtrement de mécanismes syntaxiques ne change rien à l'affaire :

[...] certains trucs sont élégants et font appel à des principes profonds d'organisation mais finalement tout ce que l'on peut espérer produire sont des systèmes qui semblent distinguer des significations en distinguant en fait des choses qui co-varient de manière fiable avec des significations. [Dennet, 1990: 86]

Dans les langages formels la référence entre les expressions et les objets est directe, alors que dans les langues naturelles, les signes ont une signification distincte de leur référence, cette dernière se trouve médiatisée par le processus de la communication qui implique un destinataire, un destinataire ayant chacun leur culture, leurs attentes et leurs représentations particulières, etc.

Par ailleurs, dans cette course à la simulation du raisonnement humain, les forces propres aux automates par rapport à la performance humaine sont oubliées. Les automates effectuent un traitement de l'information qui, bien qu'uniquement syntaxique, est fiable et efficace. Nous soutenons que les systèmes experts, tout comme les langages de programmation en général n'imposent pas en eux-mêmes une philosophie particulière quant à leur utilisation. Il s'agit de méthodologies computationnelles diverses quant à leur modalité, mais qui ont toutes en commun de contrôler l'organisation des instructions accomplies par l'ordinateur par des bifurcations conditionnelles qui permettent de choisir à un moment donné l'une de plusieurs possibilités, en fonction de la valeur d'un résultat calculé ou obtenu à une étape antérieure. Il s'agit là de l'essence du calcul symbolique. Dans les sections qui suivent, nous proposons le recours à la technique des systèmes experts, non pas pour simuler mais assister l'interprétation des textes juridiques en accomplissant des tâches où l'automate est plus performant que l'humain.

### 3 Le processus interprétatif

Nous avons vu que dans les deux cadres d'utilisation de l'ordinateur, la recherche d'information dans les bases de données textuelles juridiques de même que la modélisation des connaissances, peu de supports informatiques ne sont offerts pour l'interprétation qui joue pourtant un rôle prépondérant. Dans le premier cas, elle doit être faite partiellement par l'utilisateur, autant pour formuler une requête à l'aide de mots-clés pertinents que pour discriminer, parmi tous les documents repérés, ceux qui contiennent réellement l'information recherchée. Dans le deuxième cas, la normalisation des textes juridiques, sous la forme de règles de production ou de clauses, déplace l'objet de la modélisation de l'interprétation juridique vers le contenu des textes qui sont interprétés. Ce déplacement a pour effet d'impliciter l'interprétation juridique de telle sorte que la base de règles obtenue ne sera jamais équivalente au texte, mais une interprétation de ce dernier effectuée lors de l'ingénierie de la connaissance. L'objectif de la présente contribution est de revenir à une modélisation de l'interprétation juridique même.

Opération cognitive fondamentale, l'interprétation pourrait être définie comme un processus intégrateur qui permet de générer des significations à partir de données perceptuelles élémentaires [Meunier, 1993]. De plus, comme on interprète toujours pour une fin donnée et à partir d'un corpus de textes donné, il n'y a donc pas de mesure absolue pour en évaluer le résultat. Étant donné cette définition, l'interprétation ne peut être réduite à la connaissance et à la validation des conditions de vérité des énoncés.

Par ailleurs, des expérimentations sur le terrain ont démontré que le raisonnement humain était hautement inconsistant, autant pour la catégorisation que pour l'interprétation des configurations. À cet effet, Dennet [1990: 72] rapporte des recherches en psychologie sociale cognitive qui suggèrent que nous ne sommes que minimalement rationnels, étonnamment prompts à en arriver à des

conclusions ou à nous laisser influencer par des traits logiquement non pertinents des situations.

C'est ainsi que le recours à l'assistance de l'ordinateur trouve sa pertinence afin de fournir constance, consistance et exhaustivité. Toutefois, dans quelle mesure l'interprétation cognitive peut-elle être décomposée en opérations élémentaires sur des symboles et reproduite à volonté? Dans quelle mesure peut-on demeurer dans un paradigme calculatoire sans évacuer la marge d'indétermination, les zones de flou et les erreurs inévitables engendrées par les circonstances mêmes de l'interprétation? Telles sont les questions dont les réponses sont cruciales pour situer la place et l'intérêt du recours à l'ordinateur.

Pour fournir une assistance computationnelle pertinente à un utilisateur dans l'accomplissement d'une tâche d'interprétation, nous devons auparavant établir un modèle de cette opération. Dans une première étape, il nous semble opportun de sélectionner un cadre épistémologique pour réaliser cette modélisation. Pour ce faire, nous allons tenter de rattacher l'interprétation à un des grands types de raisonnement : inductif, déductif ou abductif. L'induction infère quelque chose de différent que ce qui est observé; ce raisonnement qui va des cas particuliers à la règle permet ainsi en quelque sorte de découvrir des lois ou régularités. La déduction s'inscrit dans un processus de démonstration où on conclut des propositions prises pour prémisses, déjà démontrées ou autopoises, d'autres propositions. Quant à l'abduction, rappelons que c'est Pierce qui a introduit l'abduction en épistémologie; il la définit comme

... une méthode pour former une prédiction générale sans assurance positive qu'elle réussira dans un cas particulier ou d'ordinaire, sa justification étant qu'elle est le seul espoir possible de régler rationnellement notre conduite future, et que l'induction fondée sur l'expérience passée nous encourage fort à espérer qu'elle réussira. [Pierce, 1867: 2.270]

L'abduction qui est présentée peut être vue comme une synthèse ordonnée des deux autres

types de raisonnements qui prend en compte le décalage entre les objets et les règles que l'on pourrait en tirer. Il suggère les hypothèses ou idées générales que la déduction développe et que l'induction met à l'épreuve [Deledalle, 1990].

Le cadre épistémologique retenu conditionne en grande partie l'implantation computationnelle du modèle. Ainsi, par exemple, si un cadre déductif semble adéquat, le problème à résoudre sera représenté par un espace d'états et la recherche d'une solution qui se fera à l'intérieur de cet espace de recherche est guidée par un but; le mode de représentation sera la logique des prédicats du premier ordre. Le raisonnement de type abductif nous semble le plus conforme à la définition cognitive de l'interprétation que nous avons retenue précédemment. Nous allons tenter de justifier le choix du raisonnement abductif et d'en déterminer l'implantation computationnelle.

L'interprétation est un problème, essentiellement de nature qualitative, relié au pouvoir créatif de l'esprit qui cherche à établir des relations entre des symboles, saisis séparément, pour identifier un schème et tirer de ce schème une ou plusieurs significations. Nous distinguons donc deux temps dans l'interprétation, la découverte du schème et l'assignation d'une signification à celui-ci. Pour identifier les schèmes (type), qui répondent potentiellement aux motivations de base du projet interprétatif, il faut trouver une configuration particulière d'indices qui prenne en compte l'ensemble des formulations des occurrences (token) d'apparition dans un corpus donné. L'identification d'un tel schème est assimilable à la formulation d'une hypothèse en ce que l'application de règles s'avère inadéquate, car il s'agit d'un processus heuristique de construction artificielle où certains symboles sont privilégiés et d'autres considérés comme du bruit afin d'isoler des régularités.

S'il est impossible de décomposer un processus heuristique en une suite de règles, il est toutefois possible de décomposer ce dernier en étapes. La première étape consiste à isoler puis à organiser et codifier le domaine des phénomènes pertinents. Dans le cas des

textes, il s'agit des contextes d'occurrences de termes qui, a priori, nous semblent pertinents en regard de l'objectif de l'interprétation. Les contextes retenus sont alors regroupés en vertu des régularités qu'ils présentent. Les groupes obtenus constituent en quelque sorte des métacontextes qui, suite à une décontextualisation et à une formalisation deviendront des ensembles de relations logiques explicites entre des variables. Ces variables représentent, en fait, des paradigmes, c'est-à-dire des séries de termes différents mais qui tiennent dans les contextes des rôles équivalents. Pour ordonner, entre eux, les termes du paradigme, différentes échelles seront découpées et, pour chacune d'elles, des gradations seront établies. Ces échelles et leurs gradations seront exploitées lors de l'attribution de la signification. Lors de l'étape de validation, ces variables deviendront des indices dans la mesure de leur présence effective dans les textes et non pas de leur projection par un lecteur en vertu de son expérience ou de son expertise. C'est donc sur le mode de l'émergence qu'apparaissent des règles de concaténation des indices en des schèmes.

Par la suite, les schèmes mis à jour doivent être testés sur l'ensemble du corpus afin d'un éprouver la validité et la couverture. Cette validation se fait par le biais d'une catégorisation à partir des paradigmes précédemment constitués. Cette catégorisation consiste à marquer les termes du texte du nom de l'indice qu'ils représentent. Un calcul qualitatif aura lieu sur ces indices potentiels pour dépister, par cumul ou agrégation, les schèmes pertinents. Ce calcul est constitué d'opérations élémentaires qui sont : le tri, le classement, la combinaison, la comparaison, la substitution, la transcodification (traduction d'un code à un autre), le repérage par appariement, etc. Ces opérations, appliquées aux objets sur lesquels on opère, termes ou indices, dans un ordre déterminé, sont répétées un très grand nombre de fois. Il est à noter qu'autant les indices que les schèmes présentent la propriété de la compositionnalité, ce qui autorise une certaine réalisation de la complexité.

Pour ce qui est de l'assignation d'une signification au schème, étant donné que

l'automate ne peut qu'effectuer un traitement syntaxique de l'information, l'ordinateur n'est que peu utile. Il ne peut guère fournir une assistance autre que de gérer et d'appliquer une table de correspondance entre des schèmes dépistés à l'aide des indices et des significations qui auront été préalablement inscrites dans la table. L'accès aux significations est directe, car l'association qualitative est difficile à implanter dans un ordinateur.

En somme, l'ordinateur peut fournir une assistance appréciable pour l'interprétation dans les phases de recueil des contextes. Une simple opération de concordance (KWIC key word in context) permet de repérer exhaustivement les contextes qui fourniront la matière première pour constituer les paradigmes qui, après validation et décontextualisation, deviendront les schèmes. Les schèmes seront par la suite décomposés en indices qui seront projetés sur les textes. Il s'agit là d'un dispositif grossier certes, mais qui pourra aider à discriminer des configurations signifiantes malgré la diversité et les fluctuations dans les formulations. C'est ce dispositif que nous avons appelé la lecture experte dont la présentation succincte sera donnée dans la prochaine section.

## 4. La lecture experte

### 4.1 LA LECTURE EST UNE EXPERTISE

Dans la foulée de précédents travaux, projets de recherches ou projets pilotes dans l'administration publique, la lecture, en tant que processus cognitif individuel et collectif, apparaît comme une expertise. Une expertise c'est la maîtrise par un individu de l'accomplissement d'une tâche dont les fondements sont la plupart du temps implicites.

«De manière générale, toute personne pratiquant intensément une activité acquiert des automatismes. Cette automatisation apporte une plus grande efficacité à l'expert, qui agit alors plus au niveau réflexe qu'au niveau conscient(...). Mais l'automatisation limite aussi l'expert lors d'un accès conscient à l'activité. Ainsi, l'expert a du

mal à expliquer ce qu'il fait (il possède un «savoir-faire» mais n'a pas de disposition pour le «faire savoir»). [Gallouin, 1988: 159].

L'expert se différencie du non-expert en ce qu'il possède une représentation interne, schématique de la tâche qui lui permet de discriminer ce qui est important de ce qui ne l'est pas et en même temps qui lui permet de voir un grand nombre de détails. Là où les autres ne voient que des cas particuliers sans liens, l'expert distingue des patrons et des structures.

Le propre de l'expertise des lecteurs est de mettre en oeuvre des stratégies efficaces de décodage et d'interprétation du «contenu» des textes afin d'en extraire de l'information pertinente. Les critères et motivations qui établissent cette pertinence sont toujours extérieurs au processus de lecture, mais jouent toutefois un rôle important dans les stratégies mises en oeuvre. En effet, les lecteurs accomplissent, avec l'information obtenue, des tâches diversifiées, telles : l'évaluation d'un dossier, l'attribution d'un contrat, etc. Dans le cas qui nous occupe, les juristes vont lire les textes juridiques pour émettre un avis légal, pour déterminer la portée d'un concept, pour extraire une règle de droit, etc.

Claude Thomasset, dans son texte sur *La lecture des textes juridiques* [1992], isole cinq niveaux de lecture : lexicale (le sens commun), conceptuelle (repérage des concepts juridiques), intertextuelle (le réseau des contextes d'occurrences pertinents des concepts juridiques), l'analyse des principes généraux qui permettent une interprétation conforme aux intentions de l'auteur, et enfin l'analyse symbolique «liée à la conception de l'être humain dans nos sociétés occidentales et aux valeurs économiques qui animent le système économique et de consommation». Le modèle computationnel de la «lecture experte» fournit support et assistance pour l'accomplissement des trois premiers niveaux. Quant aux deux autres, ils relèvent de l'aspect signification qui est laissé de côté dans le cadre de la présente contribution.

Ce terme de «lecture experte» désigne les processus de décodage et d'interprétation du «contenu» effectués par un expert du domaine

avec l'assistance d'un système informatique. La conception d'un tel système exige de concilier la méthodologie établie de l'analyse des textes par ordinateur (ATO) [Bardin, 1977] avec les données empiriques tirées de l'observation de lecteurs. C'est ainsi que les opérations fondamentales seront quelque peu réaménagées.

#### 4.2 LES OPÉRATIONS DE L'ATO

Le processus analytique est habituellement conçu en quatre opérations successives : le découpage des unités significatives du texte, la description de ces unités, l'extraction d'informations à partir des descriptions et l'interprétation des informations extraites. On arrive généralement à effectuer au moyen de l'ordinateur, à des degrés divers d'automatisme et avec plus ou moins de succès, les trois premières étapes. Cependant, à notre connaissance, aucun système basé sur l'un et/ou l'autre des trois modèles de traitement suivants : statistique [Lebart et Salem, 1988], linguistique [Coulon et Kayser, 1986] ou associationniste [Cotrell et Small, 1983; Waltz et Pollack, 1985], ne s'est vraiment attaqué à l'interprétation qui pourtant constitue l'étape cruciale de l'analyse. L'apparente impuissance de ces paradigmes s'explique en partie par la nature subjective de l'opération cognitive de l'interprétation. Elle s'explique aussi par la stratégie «étapiste» qui est habituellement mise en oeuvre dans la conception et le développement des systèmes d'ATO. Étant donné les difficultés rencontrées à chacune des étapes, celles-ci deviennent des finalités.

Le découpage du texte en mots - chaînes de caractères délimitées par des blancs - ne donne pas pour autant des unités significatives. Celles-ci sont souvent composées de plusieurs mots et leur délimitation peut donner lieu à des traitements élaborés. Ainsi, par exemple, les expressions «traitement de textes» et «assemblée générale» sont des multi-termes, aussi appelés dans d'autres contextes «segments répétés», «synapsies» ou encore «unités polylexicales». Des traitements plus ou moins sophistiqués ont été mis à contribution pour leur dépistage [Paquin, 1991; David et Plante, 1991]. En dernière analyse, le problème s'avère parfois insoluble

sans le recours à une connaissance étendue du domaine de référence [Bertrand-Gastaldy, 1990].

La description linguistique des unités et de leurs relations à l'intérieur de la phrase est elle-même l'objet d'une stratégie «étapiste» dont chacune des étapes constitue en elle-même un vaste champ de recherches. Selon cette stratégie, une série de descriptions linguistiques doivent successivement être réalisées à partir du résultat des précédentes. D'abord, on procède à une description morphologique hors contexte de chacun des mots à l'aide d'un dictionnaire. Dans le cas où plusieurs morphologies sont possibles, une description en contexte est nécessaire pour déterminer, à l'aide des mots voisins, laquelle des formes possibles est effectivement réalisée. Puis, une description syntaxique vient préciser les relations qu'entretiennent les mots à l'intérieur des propositions et les relations entre les propositions à l'intérieur de la phrase [Coulon et Kayser, 1986:107].

Une telle stratégie repose sur les postulats qu'un texte est composé d'un ensemble de structures dont les principes d'organisation peuvent être explicités et que, quoique interliées, ces structures sont suffisamment distinctes pour être décrites les unes après les autres. Or rien n'est moins sûr. La désambiguïsation d'une catégorie morphologique exige parfois une analyse syntaxique complète ou encore le recours à un contexte étendu.

Par ailleurs, la description syntaxique d'un texte, même si elle permet à terme d'accéder à une certaine sémantique relationnelle des unités, ne contribue qu'en partie à la description des niveaux supérieurs d'organisation du texte, telle la structure argumentative ou encore les «figures» [Dubois 1970], dont la prise en compte est pourtant essentielle à l'interprétation du texte. C'est qu'un texte n'est pas seulement un fait de langue mais aussi de discours. Ce dernier aspect fait appel à des connaissances externes au texte analysé, celle des autres textes et celle des conventions sociales qui en régissent la production [Foucault, 1969]. Voilà pourquoi l'étiquetage contextuel de catégories autres

que linguistiques aux unités significatives est si difficile à réaliser à l'aide de l'ordinateur.

Plutôt que de réduire la complexité du processus analytique en le découpant en étapes et que de consacrer des efforts théoriques et computationnels à la résolution automatique de tous les problèmes qui se posent pour chacune des étapes, nous proposons de focaliser sur les besoins et les pratiques de ceux qui effectuent l'analyse des textes, les lecteurs. Au lieu de chercher à mettre au point un algorithme général d'ATO, nous mettons l'accent sur la modélisation d'une lecture particulière des textes. Cette lecture consiste en grande à déchiffrer les schèmes interprétatifs dont nous avons parlé lors de la section sur l'interprétation et qui sont obtenus par un raisonnement abductif. Ces considérations nous ont amené à proposer des réaménagements aux opérations de description des textes et d'extraction de l'information pertinente.

#### 4.3 LE DÉCHIFFRAGE DE SCHEMES INTERPRÉTATIFS

La description consécutive à la segmentation sera beaucoup plus superficielle puisque tous les mots d'un texte ne sont que rarement pris en compte par le lecteur. La plupart du temps, un filtrage plus ou moins sévère est effectué en vertu de ses objectifs, de sa connaissance du domaine de référence, etc. Ainsi, la lecture se trouvera d'autant accélérée que le filtrage sera sévère. Les analyses contextuelles pour la désambiguïsation des termes auxquels des catégories multiples auront été attribuées ne seront effectuées, lors de l'étape de l'extraction, que si les occurrences sont conservées. Cette ambiguïté peut être aussi bien morphologique (par ex.: le mot «été» qui peut être un nom commun ou un participe passé) que sémantique (par ex.: le terme église peut aussi bien désigner l'institution que le bâtiment où se déroule le culte). La résolution automatique de ces ambiguïtés, qui exige un investissement considérable pour déterminer l'ensemble des règles contextuelles nécessaires, constitue un des facteurs d'empêchement dans le développement des systèmes d'ATO. Dans le cadre de la lecture experte, la description permet essentiellement

de réduire sensiblement le nombre de formes en présence et, par conséquent, le nombre de configurations différentes qui mènent à un schème interprétatif donné.

De plus, dans le cadre de la lecture experte, l'opération d'extraction de l'information pertinente à partir de la description sera revue en profondeur. Dans le cadre traditionnel de l'ATO, l'extraction se fait par un patron de fouille qui peut être arbitrairement complexe. Selon le système utilisé, le langage d'interrogation permettra la troncature et les masques sur les chaînes de caractères, il permettra aussi de tester les valeurs de propriétés déposées lors de la description et de contraindre l'adjacence. Toutefois, dans tous les cas, les requêtes sont booléennes dans la mesure où l'ensemble des contraintes doivent être réalisées pour que l'information soit extraite.

Toutefois, ce n'est pas tout à fait ainsi que le lecteur fonctionne, car, même sans la co-présence de tous les indices, celui-ci peut quand même accéder à l'information du texte. Son expertise lui permet de «lire entre les lignes», de combler l'absence d'indices ou encore de discriminer la relation probable entre les indices lorsque les configurations s'avèrent ambiguës. C'est ainsi que, plutôt que de parler d'extraction, on parlera de déchiffrement. Déchiffrer c'est aussi prédire dans la mesure où le lecteur utilise des configurations types d'indices déjà relevés précédemment au cours de la lecture ou lors de lectures précédentes. Dans le cadre des théories cognitives, on pourrait dire que le lecteur a en tête des «prototypes», les schèmes interprétatifs, réalisés dans les textes dans un état plus ou moins complet, qu'il cherche à retrouver [Taylor, 1989]. Avant de développer un exemple de lecture experte appliqué aux textes juridiques, quelques mots sur la façon dont la lecture experte est implantée.

L'assignation d'une signification au schème interprétatif est aussi vue comme une mise en relation soit avec d'autres indices ou avec des éléments externes appartenant au monde de référence ou encore à la culture du lecteur.

#### 4.4 L'IMPLANTATION DE LA LECTURE EXPERTE

L'élaboration des schèmes interprétatifs, de même que les opérations de segmentation et de description des unités pertinentes des textes se font à l'aide du Système d'Analyse de Texte par Ordinateur SATO [Daoust, 1992]. Quant à l'opération de déchiffrement, elle est plutôt effectuée par un système à base de connaissances (SBC). Nous avons déjà fait mention (section 3) de l'utilité de travailler sur les contextes d'occurrences pour élaborer les schèmes interprétatifs; SATO offre des fonctionnalités intéressantes pour produire interactivement les concordances requises sur de très grands corpus de textes et ce, avec un micro-ordinateur.

SATO offre, en plus, des fonctionnalités très efficaces pour décrire les unités pertinentes du texte. Étant donné qu'il n'y a pas une mais de multiples lectures possibles des textes, les descriptions seront variées. Celles-ci seront déterminées lors de la mise au point des schèmes interprétatifs et pourront être hors contexte ou en contexte. Si la description porte sur les mots, elle sera consignée dans des dictionnaires qui seront projetés sur le lexique des textes. Par contre, si la description porte sur des expressions complexes, elle sera assignée par le biais de concordances. Les concordances pourront aussi être utilisées pour implanter des règles de désambiguation. Ainsi, par exemple, si l'on retrouve le terme «membre» dans le voisinage immédiat du terme «église» on est sûr qu'il s'agit de l'institution et non pas de l'édifice.

Quant à la segmentation différente des mots, elle pourra être faite, soit par des concordances suite à une description morphologique (nom commun + préposition + nom commun), soit par des logiciels développés à cet effet dont TERMINO™ [David et Plante, 1991].

L'opération du déchiffrement des schèmes interprétatifs est formalisée en règles de productions pour former un SBC; les unités pertinentes des textes, regroupées et décrites avec SATO, deviennent des faits. Il s'agit là

d'une approche qui tranche radicalement avec la pratique usuelle en ATO qui est d'utiliser un langage d'interrogation pour formuler des patrons. Le générateur de SBC utilisé est celui incorporé dans l'Atelier Cognitif et Textuel (ACTE) [Paquin et Daoust, 1993].

Une telle implantation présente les avantages suivants. La complexité du dispositif computationnel requis se trouve diminuée en termes d'apprentissage du système de sorte qu'une initiation à la programmation n'est donc pas requise. Comme le système est développé par les utilisateurs eux-mêmes, la prise en compte des objectifs reliés à leur tâche est assurée alors que ces objectifs sont habituellement négligés sinon méconnus par les systèmes d'analyse de texte. De plus, une approche par SBC permet un développement modulaire des systèmes. On commence par un prototype aux dimensions modestes, quelques dizaines de règles qui sont immédiatement testées sur des données d'un corpus d'apprentissage afin de valider autant les schémas d'interprétation que les listes d'indices constituées pour leur déchiffrement. Puis, graduellement, le prototype est étendu au domaine entier. Par ailleurs, un gros volume de textes peut être l'objet d'un traitement fin sans dégradation des performances du dispositif.

Qui plus est, les SBC nous permettent d'implanter cette capacité de prédiction du lecteur qui lui permet de déchiffrer l'information, même en l'absence d'indices qui sont pourtant nécessaires. Le mécanisme du cumul de coefficients numériques est utilisé [Buchanan et Shortliffe, 1984]. Le moteur d'inférences propage l'incertitude ou encore la confiance tout au long de la chaîne inférentielle. Les formules de cumul, selon la conjoncture, provoquent, soit une atténuation, soit un renforcement du coefficient d'un fait donné. Un renforcement aura lieu lorsqu'un fait est inféré plus d'une fois, c'est-à-dire, dans le cadre de la lecture experte, lorsque plus d'un indice menant à un schème interprétatif donné auront été dépistés.

D'une part, une configuration particulière d'indices peut être associée à un schème donné malgré l'absence de certains indices, pourtant jugés nécessaires lors de son élaboration. Le

système fera quand même l'association mais avec un coefficient de certitude moindre qui, à son tour, provoquera une atténuation de la certitude lors de l'association subséquente du schème avec la signification correspondante. L'utilisation des coefficients permet un dépassement du cadre strict de la logique booléenne. En effet, comme le modèle prévoyait associer un schéma interprétatif à une conjonction d'indices, si l'un d'eux venait à manquer, sans le recours aux coefficients, l'association n'aurait pu être établie. D'autre part, les indices sont associés individuellement à un ou, le cas échéant, à plusieurs schèmes interprétatifs avec un faible coefficient qui sera renforcé au fur et à mesure du déclenchement des règles filtrant les autres indices. Ce dispositif permet une réduction de la complexité des règles nécessaires pour implanter un modèle de lecture donné en dépassant l'unicité pour déboucher sur des associations plurielles mais différenciées.

La mise en relation du schème interprétatif avec la signification correspondante est elle aussi implantée dans le SBC. En fait, la distinction entre le schème interprétatif et la signification est tenue à ceci près que dans le premier cas la relation est plus objective.

En somme, la lecture experte pourrait être résumée dans la chaîne de traitements suivante. D'abord les regroupements pertinents d'unités sont opérés à l'aide de SATO par des patrons de fouille sur des chaînes de caractères ou sur des catégories morphologiques préalablement projetées au lexique et en partie désambiguïsées. Ensuite, les mots et segments pertinents du texte sont transformés en indices suite à une description effectuée par SATO manuellement, par la projection sur le lexique de dictionnaires, ou par en contexte par des patrons de fouille. Cette description peut être plurielle : sémantique, actantielle, distributionnelle, etc. Puis, une à une, des entités logiques du texte, telles les paragraphes, les articles, les chapitres, etc. sont constituées en faits de départ et soumis au SBC. Pour chacune des entités, une chaîne inférentielle est initiée. Les règles qui testent les indices en présence sont une à une validées et, lorsque l'ensemble des indices testés sont présents dans la base de faits, les schèmes interprétatifs sont «inférés»,

c'est-à-dire qu'ils acquièrent le statut de «faits». Au cycle d'inférences suivant, les règles qui testent les schèmes interprétatifs affirmés sont validées et, le cas échéant, des significations sont à leur tour «inférées» et le tout est consigné dans un «rapport» qui peut être imprimé et conservé pour traitement statistique ou encore pour enrichir le système

étant donné les chaînages inférentiels les plus fréquents. Plus tard, les éléments du rapport peuvent, à leur tour, constituer les indices d'une méta-analyse et ainsi de suite.

La schématisation suivante représente la chaîne de traitement de la lecture experte :

## 5. Un exemple : la bonne foi

Cette section a pour unique objectif d'illustrer les notions de schème interprétatif et de «lecture experte» dans le domaine juridique. Toutefois, contrairement aux situations réelles de développement, le travail n'est pas supervisé par des juristes. Le thème retenu pour l'illustration est celui de la définition du concept flou de «bonne foi» dans le contexte de la reprise de possession d'un logement par le locateur. Le corpus de textes utilisé comporte quarante six décisions qui comptent en moyenne une page chacune. Une partie de ces décisions ont fait l'objet d'une publication alors que les autres ont été fournies par la Régie du Logement et saisies par nos soins.

La détermination d'un schème interprétatif se fait à partir de contextes d'occurrences de termes dénotant le concept-clé, ici la «bonne foi». Contrairement à une autre expérimentation qui avait été menée pour le concept de «bâtiment» dans un corpus de règlements de zonage de la Ville de Montréal [Paquin et Dupuy, 1991] où un contexte de l'envergure de la phrase suffisait généralement pour cerner le concept, il est impératif pour le concept de «bonne foi» d'élargir le contexte à tout le texte de la décision. La raison de cette différence réside en grande partie dans le statut de l'objet désigné par ces concepts.

Le «bâtiment» est une entité physique, et, à ce titre, est appréhendé par ses caractéristiques ou traits saillants; le concept est donc principalement élaboré par des compléments du nom (par ex.: *la largeur du bâtiment, l'alignement frontal du bâtiment, etc.*) qui lui fournissent en quelque sorte une substance, des ancrages dans l'espace et le temps, une

finalité, etc. La «bonne foi» est une qualité abstraite qui ne fait pas l'objet d'une description, mais d'une démonstration ou d'une preuve. En effet, dans le voisinage immédiat du terme «bonne foi» on ne retrouve que rarement des déterminants ou «spécifieurs» qui construisent le concept. Soit que le concept est inclus dans un cadre prescriptif :

Le Code civil à l'article 1659.3 impose au locateur le fardeau de démontrer "qu'il est de bonne foi, qu'il entend réellement reprendre possession du logement pour la fin mentionnée dans l'avis et qu'il ne s'agit pas d'un prétexte pour atteindre d'autres fins".

soit que le concept est inclus dans un cadre délibératif:

... la Régie est d'opinion que le locateur a établi sa bonne foi ...  
En défense, la locataire doute sérieusement de la bonne foi du locateur...

La présence ou l'absence de «bonne foi» chez un individu ne peut être directement démontrée ou prouvée; le recours à la description d'agissements est essentielle. C'est ainsi que la plupart du temps, ce concept est construit de façon négative par la narration d'une ou plusieurs situations attestées de manquement. Cette narration est composée en partie de verbes ou d'actions nominalisées qui désignent des agissements et comportements reliés à la mauvaise foi:

- accrochage
- déloger
- évincer
- menace
- règlement de compte
- subterfuge
- vengeance
- contourner
- dire des bêtises
- harcèlement
- prétexte
- tracasseries

La narration est aussi en partie composée de noms qui qualifient le climat entourant la reprise de possession:

- agressivité
- animosité
- conflit
- dispute
- litige
- mauvaises relations
- relation tendue
- relations difficiles
- tension

D'autres termes enfin désignent les motivations inacceptables pour la reprise de possession :

- hausse de loyer
- expulsion
- subdivision
- vente

Le schème interprétatif relié au concept de «bonne foi» pourrait, si l'on accepte l'analyse qui précède, ressembler à la formule suivante. Dans un contexte de reprise de possession, le locateur a eu des agissements qui ont engendré un climat désagréable pour le locataire qui manifestent des motivations inacceptables eu égard à l'article 1659.3 du Code civil et à son interprétation dans la jurisprudence et la doctrine :

REPRISE DE POSSESSION +  
 AGISSEMENTS DU LOCATEUR +  
 CIRCONSTANCES +  
 MOTIVATIONS INACCEPTABLES  
 ⇨ MAUVAISE FOI

Dans ce cas, implanter la lecture experte consistera à concevoir des dictionnaires pour la catégorisation des agissements du locateur et des motivations inacceptables. Dans le cas où il est question de reprise de possession, les dictionnaires sont appliqués et une catégorisation est effectuée. Voici un exemple d'une décision catégorisée:

Lors de l'audition, les parties sont présentes. Il est admis un bail à durée indéterminée au coût mensuel de \$215.00.

Il est prouvé un avis dans la forme et les délais prescrits par la loi pour la reprise de possession.

Le locateur met en preuve qu'il entend reprendre possession de ce logement pour l'habiter occasionnellement lui-même. [...]

La locataire explique qu'elle craint que le locateur veuille simplement vendre **\*motif=vente** la maison

en évinçant **\*agissement=éviction** la locataire. Elle déclare que les relations avec le locateur sont difficiles **\*climat=mauvaises relations** au niveau des réparations et qu'il l'a souvent menacée **\*agissement=menace** d'expulsion **\*motif=expulsion**.

Le raisonnement qui sera tenu à l'aide du système expert sera négatif. Plus il y aura d'indices, c'est-à-dire de termes du textes qui auront reçu des catégories, plus la mauvaise foi sera inférée. La bonne foi sera, par contre, inférée en l'absence d'indices.

## 6. Conclusions

Dans la présente contribution nous avons critiqué l'absence de prise en compte de l'interprétation lors du recours à l'ordinateur pour le traitement des textes juridiques, autant dans une perspective documentaire où l'accent est mis sur l'indexation des textes en vue d'un repérage conceptuel des textes, que dans une perspective d'intelligence artificielle où l'accent est plutôt mis sur la représentation des connaissances juridiques et la simulation du raisonnement juridique.

Comme solution, nous avons proposé d'enrichir les pratiques documentaires et de formalisation logique d'études empiriques sur le comportement du lecteur indexeur ou modélisateur face aux textes. L'étude de la performance humaine et des stratégies utilisées lors de l'inévitable interprétation nous apparaît centrale. Cette étude est validée par des simulations des manipulations relationnelles de symboles à partir des textes au moyen de la technologie des systèmes à base de connaissances.

Cette solution présente les avantages suivants. Elle permet une réduction des décisions méthodologiques a priori qui peuvent entrer en contradiction avec les procédures intellectuelles du lecteur. Elle exerce, de plus, une pression vers le codage uniforme des données par l'explicitation le plus à fond possible d'un grand nombre de savoir-faire et d'opérations intellectuelles qui habituellement demeurent implicites.

Elle fournit en même temps une occasion d'une réflexion épistémologique féconde sur le paradigme informatique, les systèmes experts, qui est en voie de devenir la norme de modélisation. Elle participe à définir la place qui revient à l'humain et à l'ordinateur dans le traitement cognitif de l'information tout en

sonnant le glas des systèmes formels décidables, complets et totalement cohérents.

En dernière analyse, la question de recherche derrière la lecture experte c'est de déterminer dans quelle mesure il est possible d'amener l'ordinateur à assister des traitements cognitifs reliés directement à l'interprétation des textes.

## Références

- Allen, L. E. and Saxon, C. S. 1991 "More IA Needed in AI: Interpretation Assistance for Coping With the Problem of Multiple Structural Interpretations, *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM: 53: 61.
- Ashley, K. D. 1988 "Arguing by Analogy in Law: A Case Based Model", *Analogical Reasoning*, Kluwer Academic Publishers: 205-224.
- Bardin, L. 1977 *L'Analyse de contenu*, Presses Universitaires de France.
- Bertrand-Gastaldy, S., Daoust, F., Meunier, J.-G., Pagola, G. et Paquin, L.-C. 1992 "Un prototype de système expert pour l'aide à l'analyse des jugements", *Actes du congrès Informatique et droit de l'AQDIJ*, volume C1.3.
- Bertrand-Gastaldy, S. 1990 "L'indexation assistée par ordinateur : un moyen de satisfaire les besoins collectifs et individuels des utilisateurs de bases de données textuelles dans les organisation", *ICO Québec: Gestion de l'information textuelle* 2 (3), 1990 : 71-91.
- Bing, J. 1984 *Handbook of Legal Information Retrieval*, North Holland, Part III.
- Branting, L. K. 1991 "Reasoning with Portions of Precedents", *Proceedings of the Third International Conference on Artificial Intelligence and Law*; ACM: 145-154.
- Buchanan, B. G. and Shortliffe, E. H. 1984 *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley.
- Côté, P. 1992 Approches ergonomiques des interfaces personne-machine pour la recherche d'information en ligne, *ICO, Intelligence artificielle et sciences cognitives au Québec* (4) 1-2: 9-16.
- Cotrell, G. W. et Small, S. L. 1983 "A connexionist scheme for modelling word sense disambiguation". *Cognition and Brain Theory* 6: 89-120.
- Coulon, D. et Kayser, D. 1986 "Informatique et langage naturel: présentation générale des méthodes d'interprétation des textes écrits", *Technique et science informatique* 5 (2).
- Daoust, F. 1992 *Système d'Analyse de Texte par Ordinateur version 3.6, manuel de référence*, Centre ATO•CI, Université du Québec à Montréal.
- David, S. et Plante, P. 1990 "Le progiciel TERMINO : de la nécessité d'une analyse morphosyntaxique pour le dépouillement terminologique", *Actes du Colloque international 'Les industries de la langue: Perspectives des années 1990'*, Montréal: 71-88.
- Deledalle, G. 1990 "Abduction", Auroux, S. éd. *Encyclopédie philosophique universelle vol.2 Les notions philosophiques*, Presses Universitaires de France: 4.
- Dennet, D. C. 1990 *La stratégie de l'interprète*, Gallimard trad. fr. de Engel F., *The Intentional Stance*, Bradford Book 1987.
- Dubois, J. et al. 1970 *Rhétorique générale*, Larousse.
- Fattibene, P., Martinelli, M. et Simi, M. 1989 "Automated legal reasoning using a theorem prover for modal logics" *Proceedings of the Third Conference 'Logica Informatica Diritto'* vol. 2: 357-371.
- Foucault, M. 1969 *L'Archéologie du savoir*, Gallimard.
- Gallouin, J.-F. 1988 "Systèmes experts et psychologie cognitive", *Micro-Systèmes*, décembre.
- Gardner, A. 1987 *An Artificial Intelligence Approach to Legal Reasoning*, M.I.T. Press.
- Grosz, B.J. et al. 1986 *Readings in Natural Language Processing*.
- Hofstadter, D. 1979 *Gödel, Escher, Bach : an Eternal Golden Braid*, Basic Books.
- Jones, A. J. I. 1990 "Deontic Logic and Legal Knowledge Representation", *Ratio Juris* (3) 2.
- Kowalski, A. 1991 "Case-Based Reasoning and the Deep Structure Approach to Knowledge Representation" *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM: 21-30.
- Kowalski, R. A. 1979 *Logic Programming for Problem Solving*, North Holland.
- Laferrrière, M. 1992 "Constitution d'une banque centrale de jugements des tribunaux judiciaires québécois" *Actes du congrès Informatique et droit de l'AQDIJ*, volume C1.1.
- Lebart, L. et Salem, A. 1988 *Analyse statistique des données textuelles*. Dunod.
- McCarty, L. T and Sridharan, N. S. 1981 "The Representation of an Evolving System of Legal

- Concepts: II. Prototypes and Deformations." *Proceedings IJCAI-81*.
- McCarty, L. T. 1989 "AI and Law: How to Get There from Here" *Proceedings of the International Conference on Expert Systems in Law*.
- McEnery, A. M. and Bench-Capon, T. 1989 "Battle Management for Prolog Conflict Sets, *Proceedings of the Fourth International Conference on Symbolic and Logical Computing*: 367-390.
- Meunier, J.-G. 1993 "Semiotic primitives and conceptual representation of knowledge", *Signs, Search and Communication. Semiotic Aspects of Artificial Intelligence*, Walter de Gruyter: 66-89.
- Montpetit, M. 1992 "Analyse des jugements des tribunaux judiciaires", *Actes du congrès Informatique et droit de l'AQDIJ*, volume C1.2.
- Paquin, L.-C. 1991 "Du terme au concept", *Actes du Colloque international 'Les industries de la langue: Perspectives des années 1990'*, Montréal: 313-333.
- Paquin, L.-C. 1993 *Atelier cognitif et textuel version 1.0, manuel de référence*, Centre ATO•CI, Université du Québec à Montréal.
- Paquin, L.-C. et Daoust, F. 1993 *ACTE Atelier cognitif et textuel, version 1.0, manuel de référence*, Centre ATO•CI, Université du Québec à Montréal.
- Paquin, L.-C. et Dupuy, L. 1991 "Gestion des données textuelles dans les organisations : le dictionnaire de concepts", *Actes du colloque 'Repérage de l'information textuelle'*, Hydro-Québec: 43-50.
- Pierce, C. S. 1867 *Collected Papers*, The Belknap Press of Harvard University Press.
- Poulin, D., Mackaay, E., Bratley, P. et Frémont, J. 1989 "Time Server - A Legal Time Specialist" *Proceedings of the Third Conference 'Logica Informatica Diritto'* vol 2: 733-760.
- Rastier, F. 1989 *Sens et textualité*, Hachette.
- Rastier, F. 1991 *Sémantique et recherches cognitives*, Presses Universitaires de France.
- Smith, J. C. and Deedman, C. "The Application of Expert Systems Technology to Case-Based Reasoning" *Proceedings of the First International Conference on Artificial Intelligence and Law*, ACM.
- Susskind, R. E. 1986 "Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning" *Modern Law Review* 49.
- Taylor, J. R. 1989 *Linguistic Categorization*, Oxford University Press.
- Thomasset, C. 1992 "La lecture des textes juridiques", *Lire le droit, langue, texte, cognition*, Bourcier, D. t Mackay, P. (éds), LGDJ.
- Tyree, A. 1989 *Expert Systems in Law*, Prentice Hall.
- Vossos, G.; Zeleznikow, J.; Dillon, T. and Vossos, V. 1991 "An Example of Integrating Legal Case Based Reasoning with Object-Oriented Rule-Based Systems: IKBALS II" *Proceedings of the Third International Conference on Artificial Intelligence and Law*, ACM: 31-41.
- Waltz, D. et Pollack, J. B. 1985 "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation". *Cognitive Science* 9:51-74.