

Introduction

Introduction

Le droit est un domaine où l'informatique a encore relativement peu pénétré.

L'informatique juridique s'est développée jusqu'à présent dans deux directions principales.

La première direction est celle de l'informatique documentaire, sous la forme de banques de données juridiques ou plus récemment de CD-ROM. Il s'agit essentiellement de mettre à la disposition des professionnels la masse considérable des textes juridiques et des jurisprudences selon des modes d'accès très performants que permettent les technologies modernes d'indexation.

La deuxième grande direction consiste à traduire en programme informatique des réglementations ou des jurisprudences de façon à mettre en place des systèmes qui dans un domaine spécialisé remplissent des fonctions d'aide à la formation du personnel administratif, d'aide au renseignement du public, de calcul et de liquidation d'allocation, de calcul d'impôts. Hormis la programmation procédurale classique, c'est la programmation déclarative qui est à la base des systèmes experts, qui est la plus employée dans les applications opérationnelles.

L'informatisation du droit est d'une certaine manière concomitante au développement des grosses applications de gestion. N'oublions pas que les grandes applications informatiques de gestion, dans le domaine de la gestion du personnel, de la sécurité sociale ou dans celui de la fiscalité reposent sur la programmation informatique de réglementations.

Mais ce que l'on demande à l'informatique, ce n'est pas de dérouler un programme pour produire un résultat invariant défini d'avance, mais de pouvoir interroger l'application et d'en obtenir des réponses. D'où les systèmes experts, mais aussi d'autres technologies qui font aujourd'hui l'objet de recherches actives.

Les réseaux de neurones ont par exemple été expérimentés pour essayer de prévoir l'attitude du juge face à un règlement municipal de police et d'aider les autorités municipales dans leurs décisions. Ils ont été également utilisés pour représenter les décisions d'une juridiction pour déterminer le montant d'une pension alimentaire.

Les graphes conceptuels ont été expérimentés comme interface entre des questions d'usagers et des connaissances juridiques formalisées.

La programmation orientée objet a été testée pour représenter un type de jurisprudence très sophistiqué comme la jurisprudence Dame Cachet en matière de retrait des actes administratifs, concurrentement avec la programmation procédurale et la programmation déclarative.

Enfin, la logique floue n'est pas de reste, puisque qu'elle est susceptible d'être incorporée dans des systèmes experts ou dans des méthodes se rapportant à des objets dans un programme orienté objet. La logique floue est particulièrement utile lorsque qu'il s'agit de manipuler des concepts à contenu variable et de modéliser des raisonnements impliquant un ou plusieurs concepts juridiques flous.¹

Toutes ces recherches conservent un point commun, qui est de ne jamais approcher le texte juridique directement et de nécessiter en toute hypothèse une phase d'analyse informatique au sens classique du terme.

Or, ne peut-on pas imaginer un système qui traite directement les textes normatifs et en tire des graphes conceptuels, quitte à ce que ces textes soient complétés de connaissances juridiques correspondant à la compétence juridique que possède toute personne capable de comprendre et d'interpréter un texte juridique.

Si un tel système était réalisable, nous serions en mesure de traiter de façon automatique une grande quantité de textes juridiques qui pourraient ensuite être rendus plus facilement accessibles.

On mesure les obstacles qui s'opposent a priori à une telle entreprise.

Le premier obstacle n'est pas celui qui vient le premier à l'esprit. C'est le fait que le droit n'existe réellement qu'après qu'il ait été interprété. Le texte ne suffit pas. Il faut qu'il ait été appliqué et interprété, l'interprétation étant consubstantielle à toute application du droit. Ce point de vue peut choquer les tenants, de plus en plus rares, du positivisme qui voudrait que le juge ne fasse qu'appliquer des dispositions nécessairement claires et contenues dans la loi. Mais d'abord, la règle de droit n'est pas toujours claire. En second lieu, le seul fait de décider qu'une règle est claire est déjà un acte d'interprétation. Autrement dit, l'accès au texte ne peut être pur de toute interprétation, celle-ci devant être donnée par la jurisprudence et par la doctrine. Les concepteurs de systèmes experts ont fini par en prendre conscience. L'expertise qu'ils ont traduite en règles de production est le reflet d'une interprétation déterminée de la partie du droit qu'ils ont cherché à modéliser. Enfin, la règle de droit n'est pas toujours contenue dans la loi.

¹ On consultera sur ces différents aspects les quatre ouvrages collectifs réalisés sous la direction de Danièle BOURCIER et Claude THOMASSET *Lire le droit (LGDJ, 1992)*, *Le droit saisi par l'ordinateur (Yvon Blais, 1993)*, *L'écriture du droit (Diderot Editeur, 1996)*, *Interpréter le droit : le sens, l'interprète, la machine (Bruylant, 1997)*.

Le second obstacle tient à ce que Mireille Delmas-Marty a appelé le "flou du droit" qui est dû notamment à l'importance en droit des "notions à contenu variable" selon l'expression de Chaïm Perelman.

Le troisième obstacle réside dans la complexité du droit. Modéliser un domaine restreint du droit nécessite déjà un investissement considérable. Une démarche qui ne se limiterait pas à un domaine d'application bien circonscrit semble donc a priori vouée à l'échec. Jusqu'à présent, sans parler des systèmes experts, les systèmes de compréhension automatique de texte, comme l'a souligné Catherine Fuchs (1993), restent des systèmes très limités dans leur ambition.

Une quatrième catégorie d'obstacle ne tient pas au droit lui-même, mais à la linguistique.

La conception qui a longtemps prévalu sous l'influence notamment du linguiste américain Noam Chomsky exclut du champ de la linguistique la découverte du sens, autrement dit la sémantique.

La langue est sensée être construite sur une combinatoire d'expressions bien formées en vue de produire des phrases ou des énoncés soit dans un processus de traduction, soit dans un processus de dialogue en langue naturelle. Cependant, dépourvue de toute capacité d'évaluer le sens des énoncés, la grammaire générative est impuissante à participer à un processus quelconque de représentation des connaissances à partir du texte.

Les recherches sur la modélisation du droit sont marquées par ce constat. L'impression qui en ressort est d'une part que l'on s'intéresse plus à l'application du droit à travers la jurisprudence qu'aux textes eux-mêmes qui, même s'ils ne suffisent pas à former le droit, en sont néanmoins le socle a priori incontournable; d'autre part que l'on passe du droit tel qu'il est compris, à la formalisation informatique sans autre transition qu'un processus d'analyse somme toute assez classique. Même l'utilisation de réseau de neurones pour représenter des décisions de justice nécessite à partir d'une description standardisée une phase de codage relativement lourde. Tout se passe comme si l'on prenait soin d'éviter de recourir à la linguistique, science encore trop peu stabilisée pour qu'on se hasarde à l'utiliser.

Cette attitude de réserve vis-à-vis de la linguistique a de quoi surprendre dans la mesure où la linguistique a accompli des progrès tout à fait spectaculaires et que seul le traitement linguistique nous autorise à envisager le traitement de textes en masse.

Voyons donc les éléments qui permettent de modérer le jugement pessimiste sur les possibilités d'une automatisation du traitement des textes juridiques.

Il y a d'abord une première raison qui tient aux caractéristiques propres du langage juridique.

Le langage juridique est un langage technique qui pour certaines de ses composantes est étranger au langage naturel, le sens des mots voire la syntaxe se démarquant de leur sens ordinaire en langue naturelle, et la syntaxe pouvant exceptionnellement déroger aux règles de la grammaire commune. On fera observer que le langage naturel renferme lui-même des usages multiples et discordants, et en théorie pure, on pourrait soutenir que les mots n'ont jamais le même sens selon les locuteurs qui les prononcent, ni jamais le même

sens selon les auditeurs qui les entendent. Le langage du droit ne dérogerait pas à cette règle et serait en quelque sorte une exception parmi la multitude d'exceptions qui composent le langage naturel. Cette remarque nous amène à considérer le langage du droit, plus comme sous-ensemble du langage naturel que comme un langage qui lui serait étranger et nous espérons démontrer que ce langage, sous-classe du langage naturel, obéit à des contraintes qui le rendent sensiblement plus simple que le langage naturel, qu'il s'agisse du langage parlé ou du langage littéraire.

Nous ne prendrons qu'un exemple. Une des difficultés les plus grandes du traitement automatique du langage naturel est le rôle de l'implicite. L'implicite peut être impliqué par l'énoncé lui-même. L'exemple classique "Paul a cessé de fumer" trouve sa solution dans le texte analysé. Si Paul a cessé de fumer, c'est donc qu'il fumait auparavant. Mais cet énoncé a priori très simple renferme également une ambiguïté. En effet, on peut l'interpréter comme signifiant que Paul fumait régulièrement et qu'il a mis un terme à cette habitude nuisible pour sa santé. Mais rien n'interdit de penser que Paul fumait cinq minutes auparavant et qu'il vient de cesser de fumer de manière toute provisoire. Seul le contexte de l'énonciation et la connaissance que les interlocuteurs ont de Paul permettent de lever cette ambiguïté. Alors que l'implicite peut être résolue en faisant appel à la logique interne de la phrase, l'ambiguïté nécessite de faire appel au contexte. Or, la connaissance du contexte est une donnée implicite qui fait que l'auditeur est en mesure de comprendre ce que lui dit le locuteur.

Dans un texte normatif qui est en principe général, intemporel et impersonnel, il y a peu de risque de rencontrer de telles ambiguïtés provoquées par la méconnaissance du contexte de l'énonciation. Par contre, l'implicite joue un rôle très important mais spécifique.

En premier lieu, on peut rencontrer, comme dans l'exemple précédent, des données implicites impliquées par le texte lui-même. Dans "nul ne peut être puni qu'en vertu d'une loi..." (article 8 DDHC), la notion de faute ou d'infraction est impliquée par le verbe "punir". Dans "Sont seuls punissables les faits constitutifs d'une infraction..." (article 112-1 du nouveau code pénal), ce qui est implicite, ce n'est pas la faute ou l'infraction, mais l'auteur de l'infraction. Il est difficile de punir un fait. Il est plus normal de punir la personne coupable d'un fait qui lui est reproché.

On peut rencontrer une autre forme de données implicites lorsqu'un texte définit un organisme ou une instance en s'abstenant de le ou de la créer. Si l'instance ou l'organisme n'existait pas déjà antérieurement au texte concerné, on doit en inférer que le texte qui le ou la définit est en même temps un texte institutif. Contrairement au cas précédent, ce type de donnée implicite ne peut pas être déduit directement de la logique du texte étudié, mais implique d'avoir la connaissance de la situation antérieure au texte concernant l'existence ou l'inexistence de l'organisme ou de l'instance considérée. Mais, il est important d'observer que ce n'est pas la connaissance du contexte de l'énonciation qui permet de mettre au jour la donnée implicite, mais la connaissance de l'état existant du droit à un moment donné relativement à cette question.

Ce cas n'est pas très différent de l'usage, courant dans les textes juridiques, de concepts qui renvoient à une connaissance générale du droit. Par exemple, la loi du 10 janvier 1984 sur l'enseignement supérieur indique que les universités sont des établissements publics. Le texte n'explique pas ce qu'est un établissement public, mais renvoie à des connaissances juridiques plus générales que se doit d'avoir tout praticien du droit. Ces

connaissances impliquées par les textes sont tout à fait susceptibles d'être explicitées dans la base de connaissance dont le logiciel de traitement automatique de textes normatifs paraît ne pouvoir se passer.

En fait, ce dernier point fait référence à une notion capitale dans tout processus de communication, c'est la notion de compétence linguistique et plus précisément sémantique. Si dans tout processus de communication il y a un codage du contenu de pensée par le locuteur (processus onomasiologique) suivi d'un décodage par l'auditeur ou destinataire du message (processus sémasiologique), ces deux processus en réalité totalement interactif dépendent étroitement, du point de vue de la qualité de communication et de la compréhension, de la compétence linguistique et sémantique de chacun des interlocuteurs et de la connaissance réciproque qu'ils ont de leur compétence.

Cette observation élémentaire est vraie dans la communication usuelle, mais elle l'est tout autant et de manière plus manifeste car plus sélective lorsque l'objet du message relève d'un domaine spécialisé, qu'il s'agisse d'un document scientifique ou juridique.

Le fait que l'implicite soit commandé non par le contexte de l'énonciation, mais par les connaissances générales du droit existant ou des concepts qu'il véhicule, élimine du champ de la recherche des spécifications du langage du droit un vaste domaine qui présente au contraire une très grande importance quand il s'agit de textes littéraires ou de la conversation ordinaire.

L'évocation des concepts généraux en droit, dont le contenu reste généralement à l'état implicite dans les textes normatifs, permet d'aborder une autre difficulté de la modélisation en droit qui est l'existence de concepts dits flous ou à contenu variable, tels que "intérêt général", "utilité publique", "ordre public", "urgence", "gravité", "faute lourde", etc. C'est évidemment face à ce type de concepts que la marge d'appréciation des juges est la plus importante et que se trouve posée la vraie question de l'interprétation en droit. En réalité, on doit considérer que ces concepts se construisent à mesure que la jurisprudence en fait application à des cas de plus en plus nombreux. Et ce processus de construction, et de transformation continue du droit, tend à en réduire l'incertitude et est lui-même susceptible de modélisation selon des méthodes appropriées. La connaissance autour d'un concept déterminé est ce qui donne à ce concept un sens.

Si l'on regarde maintenant du côté de la linguistique, on constate que les développements les plus féconds qu'a connus cette discipline depuis une trentaine d'années ont plus été d'ordre sémantique que d'ordre syntaxique. Et si la syntaxe a été modifiée, c'est sous l'influence de la sémantique. En fait, la sémantique s'est développée dans plusieurs directions sur lesquelles nous reviendrons plus en détail et qui sont :

- la sémantique componentielle ou micro-sémantique
- la sémantique de l'énoncé
- la pragmatique
- la logique sémantique ou logique du sens
- la sémantique textuelle

C'est au tamis de ces différentes approches, lesquelles précèdent le choix des outils de la modélisation, qu'il paraît particulièrement prometteur de soumettre les textes normatifs. L'objectif épistémologique ultime qui est visé, est qu'en cherchant à dégager et à modéliser le sens des textes normatifs, on se mette en mesure de comparer et d'établir l'équivalence de sens ou la différence de sens entre deux textes normatifs. C'est le résultat qui doit être obtenu si l'on veut atteindre l'objectif pratique consistant à utiliser la modélisation des connaissances à des fins de codification et de simplification des textes normatifs.

La thèse s'intitule : « L'apport de la modélisation des connaissances à la codification et à la simplification des textes normatifs ». Bien que cet intitulé puisse être immédiatement compris et interprété, il n'est pas inutile d'en analyser chacun des termes.

Nous passerons rapidement sur l'idée de « l'apport ». Nous entendons en effet faire la démonstration, que nous espérons convaincante, que les moyens très étendus offerts par la modélisation des connaissances dans le cadre des développements récents de l'intelligence artificielle appuyés sur les développements non moins récents, relevant des trente dernières années, dans le domaine de la linguistique et de la sémantique, peuvent rendre les plus grands services dans la compréhension, la simplification et l'accès au droit du public le plus large, des professionnels aux administrés et justiciables.

L'apport prendra donc toute sa signification à l'issue de la thèse elle-même.

La « modélisation » appelle de plus amples commentaires. L'idée même de modélisation implique que l'on en donne une définition, que l'on évoque les moyens de la modélisation, et que l'on tente de répondre à la question de savoir si, épistémologiquement, le droit est modélisable. Cette dernière question est d'ailleurs tout à fait grave, car elle pose d'abord la question fondamentale de la place de l'interprétation dans le droit et celle des rapports entre l'interprétation du droit et la modélisation du droit.

