



Lidil

Revue de linguistique et de didactique des langues

38 | 2008

Langue du droit

Du modèle documentaire et linguistique à l'outil informatique juridique : *SIRC, Système Interactif pour la Rédaction de Contrats*

Thomas Lebarbé



Édition électronique

URL : <http://journals.openedition.org/lidil/2779>

DOI : 10.4000/lidil.2779

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

Date de publication : 1 décembre 2008

Pagination : 71-88

ISBN : 978-2-84310-130-4

ISSN : 1146-6480

Référence électronique

Thomas Lebarbé, « Du modèle documentaire et linguistique à l'outil informatique juridique : *SIRC, Système Interactif pour la Rédaction de Contrats* », *Lidil* [En ligne], 38 | 2008, mis en ligne le 01 juin 2010, consulté le 19 avril 2019. URL : <http://journals.openedition.org/lidil/2779> ; DOI : 10.4000/lidil.2779

Ce document a été généré automatiquement le 19 avril 2019.

© Lidil

Du modèle documentaire et linguistique à l'outil informatique juridique : *SIRC, Système Interactif pour la Rédaction de Contrats*

Thomas Lebarbé

La rédaction de documents juridiques obéit à un certain nombre de règles fondamentales. Ces documents doivent en effet être clairs, structurés et sans ambiguïté, afin que les lecteurs puissent les interpréter de la même manière et puissent s'assurer que tous les points envisagés sont traités. Ces conventions rédactionnelles se retrouvent à plusieurs niveaux : la structure du document lui-même, la structure des clauses dont il est constitué, les structures stylistiques, et enfin les choix terminologiques. Le contrat est un exemple de ce type de document, le contrat de licence de droit de propriété intellectuelle un cas parmi tant d'autres et objet de cet exposé.

- 1 Il existe un certain nombre de travaux dont l'objectif est de normaliser les structures documentaires juridiques. Nous avons noté notamment les efforts de *Legal XML* (LXML, 2005) ou ceux de la *Global Justice Information Sharing Initiative* (JXDM, 2005) avec le développement du JXDM (*Justice XML Data Model*), à l'initiative de nombreuses entités administratives et judiciaires américaines. Cottin (2005) montre en revanche que ce genre d'initiatives est plus lent à se mettre en place au niveau français.
- 2 Le travail présenté ici est le fruit d'un projet de recherche et développement entre le laboratoire de Linguistique et didactique des langues étrangères et maternelles (LIDILEM) d'une part et le cabinet juridique Breese, Derambure, Majerowicz, d'autre part. Ce travail s'inspire des propositions des nombreux consortiums dont l'objectif est la structuration documentaire et de la diversité des solutions proposées. Un outil, nommé SIRC (Système Interactif pour la Rédaction de Contrats), a été développé dans le dessein d'aider le rédacteur de contrat dans sa tâche. Cette aide lui est apportée sous différentes formes :
 - **1. une base documentaire**, historique rédactionnel du rédacteur et de ses collègues (en l'occurrence du même cabinet juridique pour des raisons de confidentialité) ;

- **2. une interface de rédaction collaborative**, permettant une intervention extérieure telle celle d'un collègue du cabinet, d'un collègue représentant l'autre partie contractante, ou des clients eux-mêmes ;
 - **3. une base de schémas structurels**, répertoriant les structures les plus stéréotypiques en fonction d'un certain nombre de critères qualifiant le contrat (licence exclusive ou pas, par exemple) ;
 - **4. une base personnalisée**, qui permet certes au rédacteur de se fonder sur des clauses existantes mais surtout d'adapter certaines phrases ou tournures à son style et d'en garder trace exclusive ;
 - **5. un ensemble d'outils statistiques**, et plus particulièrement de statistique linguistique, permettant d'alerter le rédacteur si une clause manque ou inversement n'est pas présente usuellement, ou si des erreurs stylistiques semblent avoir été commises ;
 - **6. des outils standards du traitement automatique des langues**, tels un correcteur orthographique et un correcteur grammatical (en cours de développement) dont les principes de corrections pourront être adaptés spécifiquement au style juridique.
- 3 Dans un premier temps, nous présentons certaines observations structurelles sur le corpus de contrats de licences de brevets et de marques. Ensuite, les différentes formes d'aide sont présentées dans le même ordre que la liste ci-dessus, au sein de la seconde section. En conclusion, nous présentons certaines perspectives d'extension d'un point de vue linguistique mais aussi ergonomique et juridique.
- 4 SIRC a été développé en interface Web qui peut donc être limité à un intranet d'entreprise ou être mis à disposition d'interlocuteurs externes sur un serveur Internet. Dans tous les cas, la sécurité des informations et la limitation des accès est assurée, qu'il s'agisse d'un intranet ou pas.

1. Particularités du corpus

Le Cabinet de Conseil Juridique Breese, Derambure et Majerowicz s'est forgé une place de leader dans le domaine du conseil en droit de la propriété intellectuelle depuis de nombreuses années. Nous avons donc eu à notre disposition une large collection de contrats de licence de brevet, de modèles et de marques.

- 5 Par souci de respect de la confidentialité du conseil juridique, mais aussi par respect et protection du savoir-faire des experts, les contrats ont été rendus anonymes et seuls des extraits significatifs, mais non confidentiels, seront présentés à titre d'exemple.
- 6 L'ensemble des contrats étudiés est le fruit du travail d'un grand nombre de rédacteurs. D'autre part, ces rédacteurs ne sont pas tous issus de la même formation et n'ont pas effectué leur carrière uniquement au sein du même cabinet juridique. Par conséquent, les observations faites dans les pages qui suivent ne sont ni la description de la compétence d'un rédacteur, ni celle de la compétence rédactionnelle d'un cabinet, mais une généralisation sur la rédaction d'une forme de contrat dans un des domaines d'application du droit : la gestion de la propriété intellectuelle.

1.1. Anonymation

Rendre des documents juridiques anonymes de manière automatique est une tâche complexe qui fait l'objet de nombreux travaux de recherche (notamment dans le cadre de l'anonymation d'arrêts). Aucun, à notre connaissance, n'est en mesure de donner un

résultat fiable et efficace à moindre coût. Le travail a donc été effectué manuellement, ce qui a permis de mettre en évidence certains motifs récurrents.

- 7 En effet, chaque contrat utilise un système simple de définition de raccourcis, de termes anaphoriques. Dans la pure logique argumentaire, les définitions sont données avant l'utilisation des termes anaphoriques.
- 8 Dans le cas récurrent présenté ci-dessous (fig. 1), les parties du contrat doivent être désignées clairement, par une adresse pour les personnes physiques et par une identification commerciale et administrative pour les personnes morales. Il est conventionnel d'assigner aux parties un terme descriptif de leurs rôles respectifs (« concédant » et « licencié » dans le cas de la figure 1).
- 9 Il est à noter que de manière générale, mais pas exclusive, le terme anaphorique est marqué typographiquement afin d'en faciliter l'identification visuelle pour le lecteur, le plus souvent par une écriture en majuscule comme dans les figures suivantes.

10 **ENTRE LES SOUSSIGNÉS :**

11 Monsieur [Prénom] [NOM], demeurant [Numéro], [rue] à [Ville]
 ([CodePostal]),

12 Ci-après désigné "le **CONCÉDANT**"

13 D'une part,

14 **ET**

15 La Société [Nom Société], [Status Juridique], [Identification Commerciale],
 dont le siège social [Numéro], [rue] à [Ville] ([CodePostal]),

16 Représentée par [Monsieur] [Prénom] [Nom]

17 Agissant en qualité de [Qualité],

18 Ci-après désignée le "**LICENCIÉ**"

19 D'autre part.

Fig. 1. - Exemple de définition du terme anaphorique

- 20 Il s'agit à la fois d'une définition d'un référent, dont l'usage unique est de pointer vers l'identité complète, mais aussi d'un terme non ambigu, et d'une anticipation sur le contrat, par le choix du référent. L'expert, à la simple lecture de cette partie de l'en-tête du contrat, peut anticiper sur le contenu du contrat, en ayant une connaissance implicite de modèles structuraux dont les détails seront décrits ultérieurement.
- 21 La déictique n'est pas uniquement nécessaire à la désignation des parties. Un ensemble de termes est généralement défini et propose des raccourcis non ambigus désignant les éléments importants et récurrents du contrat. Dans l'exemple de la figure 2 ci-après, le terme « BREVETS » est défini comme une succession de descriptions administratives de brevets (nom et numéro d'enregistrement). Il est important de noter aussi que les définitions suivantes font référence à des définitions antérieures.

22 - On entendra par BREVETS au sens du présent contrat les titres de propriété
 industrielle suivants :

23 - le brevet français [nom et identification INPI du brevet]

24 - ...

- 25 - On entendra par **PRODUITS CONTRACTUELS**, les produits mettant en œuvre une au moins des revendications des **BREVETS**.
- 26 - On entendra par **CONTRAT** le présent document ainsi que les documents annexés au moment de la signature.
- 27 - On entendra par **INFORMATION CONFIDENTIELLE** toute information scientifique, technique ou commerciale divulguée par l'une des Parties à l'autre Partie, sous forme écrite, matérielle ou orale et, désignée comme confidentielle au moment de sa divulgation par la Partie divulgante.

Fig. 2. - Définitions de termes « classiques »

- 28 Lorsque les définitions ont été apportées, certaines clauses génériques à une classe de contrats (ici « licence exclusive de brevet »), sont retrouvées à l'identique dans de nombreux contrats de la même classe (voir figure 3), qu'ils portent ou non sur le même type de brevet.

- 29 **Article 1 : OBJET**
- 30 1.1 Le **CONCÉDANT** concède au **LICENCIÉ**, qui l'accepte, une licence exclusive des **BREVETS** pour la fabrication et la commercialisation des **PRODUITS CONTRACTUELS**.
- 31 1.2 Afin de faciliter la fabrication industrielle et la commercialisation en nombre des **PRODUITS CONTRACTUELS**, le **CONCÉDANT** transfère au **LICENCIÉ** le **SAVOIR-FAIRE** dont il dispose sur les **BREVETS**.

Fig. 3. - Utilisation des référents dans des clauses essentielles

- 32 Il existe aussi un système d'anaphores et d'anaphores composées. Par anaphore composée, nous entendons : un terme qui n'est pas défini explicitement, mais qui fait référence à un autre défini sous une autre forme ultérieurement dans le document. Dans ce cas, le référent est accompagné d'un modificateur qui s'applique donc au terme référé.
- 33 L'exemple le plus flagrant, mais aussi le plus utile, est la date de signature qui n'est pas définie en tant que telle, mais qui reste parfaitement explicite. Elle est en général définie par l'apposition manuelle de la date avant la signature (voir fig. 4).

- 34 Fait en neuf exemplaires, dont un pour chacune des Parties et un pour les inscriptions auprès des Offices de propriété industrielle concernés.
- 35 Fait à....., le

Fig. 4. - Identification de la date de signature

- 36 Dans ce cas particulier, le contrat fait anaphoriquement référence à cette date avant sa définition, « prend effet à sa signature », mais aussi de manière plus complexe, « à la date anniversaire de sa signature », indiquant soit un pas dans le temps, soit une régularité.

- 37 **ARTICLE 4 - DUREE – RESILIATION**
- 38 Le présent Contrat prend effet à sa signature pour une durée de 2 (deux) ans.
- 39 Au-delà de cette période, il sera renouvelé annuellement par tacite reconduction. A la date anniversaire de sa signature, il pourra être modifié par un avenant, signé des **PARTIES**, qui en précisera l'objet.

Fig. 5. - Exemple double d'usage anaphorique du terme signature

- 40 Enfin, nous avons pu observer un système de référence classique à des positions relatives ou absolues dans le la structure du document.
- 41 Notons toutefois le faible effectif des pointeurs relatifs. En effet, il est difficile de trouver des termes adéquats pour désigner les pointeurs relatifs. Seuls « suivant » et « précédent » échappent à cette règle. Mais ils restent fragiles lors d'une rédaction de contrat où l'insertion d'une clause peut déstabiliser la référence à « la clause précédente ».
- 42 **Article 1 : Objet :**
- 43 Par le présent contrat, le donneur de licence concède au licencié une licence exclusive pour le PROCESS logiciel tel que caractérisé en annexe, aux conditions ci-après.
- 44 **Article 5 : Propriété intellectuelle.**
- 45 5.1 : Le donneur de licence déclare être investi de tous les droits sur le logiciel donné en LICENCE.
- 46 Tout acte du licencié non couvert par l'article 3 serait contrefacteur et justifierait des poursuites de la part du donneur de licence.

Fig. 6. - Usage de pointeurs positionnels dans le document

- 47 Ce système de références forme un réseau, ou un ensemble de réseaux, d'ordre sémantique, qui se calque sur une structure hiérarchique. Dans le cadre d'une bonne gestion documentaire, il est nécessaire de conserver ces propriétés et d'en jouer dans l'aide à la rédaction.

1.2. Découpage structurel

La structure physique d'un contrat est relativement simple (voir figure 7) : en-tête définissant le titre du contrat, désignation des parties, ensemble de clauses numérotées, enfin datation et signature des parties. Généralement, le nombre de clauses est peu élevé, chacune traitant un point de droit particulier au contrat.

- 48 Intitulé du contrat
- 49 Désignation des parties
- 50 Clauses
- 51 Clause 1
- 52 Clause 2
- 53 Clause 3
- 54 ...
- 55 Clause n
- 56 Signature

Fig. 7. - Structure physique du contrat

- 57 Dans le cas de notre corpus d'étude, nous avons pu constater d'autres formes de régularités structurelles, mais d'ordre juridique, se superposant aux régularités structurelles de forme du contrat stéréotypique.
- 58 Nous avons ainsi pu identifier des classes de clauses et des ordres d'apparition réguliers :
- définitions,
 - opération juridique,
 - prix,
 - garanties,
 - durée du contrat,
 - fin du contrat,
 - attribution de compétence (juridiction).
- 59 Chacune de ces classes est constituée d'une clause contractuelle, ou d'un ensemble cohérent et figé de clauses contractuelles.
- 60 Enfin, en sous-catégorisant les contrats, nous avons obtenu des classes où nous retrouvons des similarités structurelles dans la forme comme dans le fonds des clauses.
- 61 Toutes ces propriétés sont les premières mises en évidence dans l'aide à la rédaction.

1.3 Structure de document

De ces observations, nous avons constitué un squelette de document, qui suit la DTD (*Document Type Definition*) suivante (voir figure 8) mais qui a été implanté sous forme d'une base de données relationnelle (voir figure 9). La DTD permet de contraindre la rédaction dans un certain nombre d'éditeurs de document structuré tandis que la base de données offre une structure plus efficace à manipuler qu'un ensemble de fichiers XML (Extended Mark-up Language) par des systèmes de recherche d'information et de suggestion de contenu (Lebarbé *et al.*, 2008).

```

62 <!ELEMENT Contrat (en-tete, parties, clause+, signature)>
63   <!ELEMENT en-tete (#PCDATA)>
64   <!ELEMENT parties (personne+)>
65     <!ELEMENT personne (civile|morale, representant?, variable)>
66       <!ELEMENT civile (nom, prenom, adresse)>
67       <!ELEMENT morale (nom, raison_sociale, adresse, identification)>
68       <!ELEMENT representant (nom, prenom, qualite)>
69         <!ELEMENT nom (#PCDATA)>
70         <!ELEMENT prenom (#PCDATA)>
71         <!ELEMENT adresse (#PCDATA)>
72         <!ELEMENT raison_sociale (#PCDATA)>
73         <!ELEMENT identification (#PCDATA)>
74         <!ELEMENT qualite (#PCDATA)>
75     <!ELEMENT clause (titre, texte)>
76     <!ATTLIST clause classe (definition | operation_juridique |

```

```

77         prix | garanties | duree | fin |
78         juridiction | CDATA) #REQUIRED>
79 <!ELEMENT titre (#PCDATA)>
80 <!ELEMENT texte (#PCDATA, pointeur*)>
81 <!ELEMENT variable (#PCDATA)>
82 <!ATTLIST variable type (declaration | pointeur) #REQUIRED>
83 <!ELEMENT signature (variable+)>

```

Fig. 8. - DTD d'un contrat typique

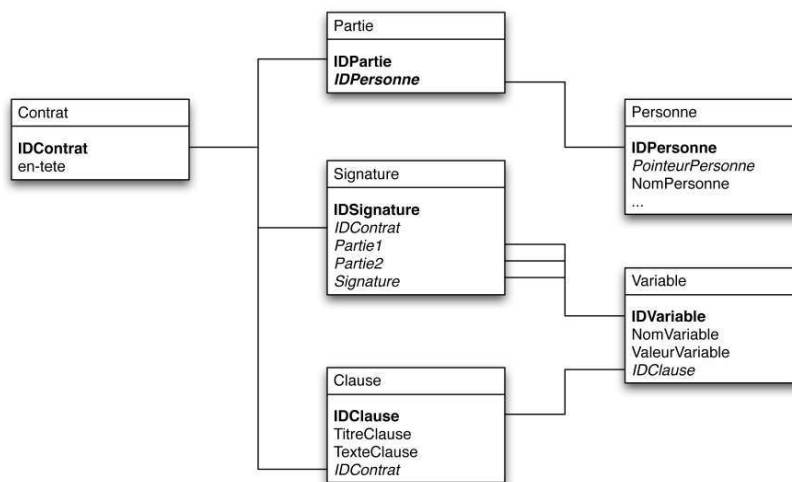


Fig. 9. - Schéma relationnel d'un contrat typique

84 L'un comme l'autre schéma peuvent être interprétés comme suit :

- Un contrat est constitué :
 - d'un en-tête,
 - d'une déclaration des parties et des raccourcis pour les désigner,
 - d'une suite ordonnée de clauses,
 - d'une signature et d'une date de référence.
- Une clause est constituée :
 - d'un numéro (calculé automatiquement) et d'un titre,
 - d'un contenu pouvant intégrer une définition d'un terme.

85 Bien que les deux modèles (document XML et base de données documentaires) soient compatibles, nous avons préféré implanter le modèle de base de données. À l'époque du développement du premier prototype, les navigateurs et modules de services compatibles XML présentaient de nombreuses faiblesses et l'intégration au système Intranet du commanditaire nécessitait l'utilisation d'un système PHP-SQL existant.

2. SIRC : un système à facettes multiples

Le système SIRC est un système informatique d'aide à la rédaction mettant en application les observations stylistiques et structurelles présentées dans la section précédente. Il s'agit d'une application logicielle complète et elle répond avant tout aux besoins du commanditaire. Dans cette section, nous présentons séparément et successivement les différentes facettes de l'application et l'influence de nos observations sur leur constitution.

2.1. Base documentaire

Dans un cabinet juridique, chaque conseiller rédige un grand nombre de contrats par année. Ces contrats sont généralement rédigés à l'aide d'un outil de traitement de texte standard et stockés sur le disque du rédacteur en fonction de ses propres critères. Il semblait impératif de pouvoir pérenniser l'ensemble des contrats conçus par le cabinet dans une structure commune, sans toutefois nuire aux spécificités de classement de chacun des rédacteurs.

- 86 Chaque contrat étant constitué de clauses fréquemment réutilisées, il paraissait utile de concevoir une base de clauses où ces dernières seraient caractérisées et par conséquent catégorisées en fonction d'un certain nombre de critères tels le type de contrat dans lequel apparaît régulièrement la clause, mais aussi la préférence de chaque rédacteur à utiliser l'une plutôt que l'autre forme d'une clause donnée ou tout simplement le rédacteur original de la clause.
- 87 Chacun de ces critères a été considéré comme une clef d'accès possible à une clause déjà rédigée afin qu'elle soit intégrée dans un contrat en cours de rédaction ; plusieurs arbres sont donc construits dynamiquement et affichés dans l'interface à cet effet.
- 88 Par ailleurs, comme nous l'avons évoqué auparavant, chaque clause de chaque contrat intégré à SIRC est un enregistrement d'une table de base de données. Par conséquent, il est possible d'interroger simplement la base de données en fonction de la présence et de l'absence de termes ou expressions au sein de leur titre ou de leur contenu – ces expressions peuvent inclure des jokers (ou *wild-cards*).
- 89 En complément de ce système de recherche de terme, il est prévu à court terme d'intégrer un lemmatiseur et un étiqueteur morpho-syntaxique. Ces outils issus des technologies du traitement automatique des langues (TAL) permettent d'associer à chaque mot d'un texte respectivement son lemme et son étiquette morpho-syntaxique (sa catégorie et sa sous-catégorie syntaxique non ambiguë). Ceci permettra de rechercher des formes d'écritures, tels les adjectifs utilisés pour qualifier un terme technique, ou les formes verbales utilisées dans tel ou tel contexte. Cette méthode s'inspire notamment des travaux de Didier Bourigault et de son équipe (Frérot *et al.*, 2003) ; une approche endogène de l'aide à la rédaction qui permet à l'utilisateur de rédiger par l'exemple des rédactions existantes.

2.2. Interface de rédaction collaborative

Un contrat n'est pas un écrit de solitaire, bien au contraire. Il est le résultat d'une négociation et d'une collaboration active entre deux parties ou plus (pour les contrats

synallagmatiques¹), et dans les cas qui nous concernent, entre deux clients et les conseillers juridiques qui les assistent.

- 90 Il était donc nécessaire d'intégrer à SIRC un aspect rédaction collaborative. SIRC propose une approche comparable au modèle des systèmes de gestion de contenu (CMS – *Content Management Systems*, Wiki) (Leuf et Cunningham, 2001). Il est ainsi possible de donner à un nombre limité d'utilisateurs la possibilité de modifier le contrat (ajout, suppression, modification de clauses), tout en gardant une trace des modifications effectuées et des auteurs desdites modifications. Le responsable de la rédaction peut alors revenir sur les modifications de chacun le cas échéant.
- 91 Cette approche a quelques conséquences notables sur le fonctionnement du système SIRC : d'une part, la notion de rédacteur principal, en l'occurrence le conseiller du cabinet juridique qui initie la rédaction du contrat ; d'autre part, l'affichage de l'aide doit à l'utilisateur connecté. Il est inconcevable qu'un client et encore plus qu'un concurrent puisse accéder à la base de contrats ou de clauses.
- 92 Il est par ailleurs en projet d'intégrer la signature numérique, la dernière étape vers le document juridique authentifié.

2.3. Schémas de rédaction

Nous avons montré précédemment que des modèles prototypiques de contrats ont pu être définis grâce à l'étude de ceux-ci, modèles structurels mais pas modèles stylistiques ni linguistiques. Il en découle qu'il existe des schémas rédactionnels, mais que les contraintes stylistiques sont moins marquées.

- 93 Nous avons donc pu produire un ensemble de schémas, de pattern-designs arborescents, donnant accès à différentes formes écrites d'une même clause. Ceci permet au rédacteur de choisir le modèle de contrat type qu'il désire rédiger, puis pour chaque clause, un style ou un autre, à moins qu'il ne désire rédiger lui-même la clause. Cela facilite la première rédaction d'un contrat tout en permettant autant de liberté que nécessaire. Simultanément, si l'utilisateur omet (délibérément ou pas) une clause ou une autre, le système le lui signale par un panneau d'avertissement dans la structure du contrat.
- 94 Il en est de même pour un certain nombre de « variables » d'environnement. Les termes pointant vers une définition sont appelés au sein des clauses à l'aide de crochets (ex. : '[BREVETS]'). Si le terme n'a pas été défini ailleurs dans le document – à l'aide d'accolades (ex. : 'on entendra par {BREVETS}...') – un panneau d'avertissement apparaît dans l'aperçu du document.

Les dates sont traitées de la même manière. La seule différence est dans la déclaration. La déclaration est faite *a posteriori*, et les pointeurs sont indirects. D'autre part, le seul intérêt de définir les pointeurs est le calcul des échéanciers du contrat, de manière à générer des avertissements pour le responsable du contrat. Il existe donc un balisage spécifique '{DATE jj-mm-aa}', qui permet de déclarer la date de signature du document. Si la chaîne 'jj-mm-aa' est inscrite telle quelle, le système ignorera la déclaration de date. En revanche si cette chaîne est remplacée par sa valeur (ex. : « 24-05-05 »), le système calculera automatiquement les échéances. Les pointeurs sont déclarés par '[DATE n/i/u]' où 'n/i/u' est une forme d'expression régulière correspondant à la description du phénomène. Le phénomène consiste toujours en un événement après une certaine période, ou une série d'événements répétés régulièrement à compter de la date de signature (une déclaration

plus complexe peut se composer des deux précédentes). Nous avons donc n ($1 \leq n \leq +\infty$) répétitions d'un phénomène, avec une périodicité de i ($1 \leq i \leq +\infty$) unités ($u \in \{j|s|m|b|t|s|a\}$), avec j pour jour, s pour semaine, m pour mois, b pour bimestre, t pour trimestre, s pour semestre, a pour année).

2.4 Diversité stylistique

Grâce à l'aide à la rédaction, et tout en bénéficiant d'une liberté rédactionnelle totale, l'utilisateur en vient à produire ses propres clauses. Celles-ci sont présentes dans la base de donnée, ne serait-ce que parce qu'elles correspondent à une réécriture d'une clause utilisée dans un contrat particulier. D'autre part, la norme simple, pour pointer vers les éléments courants, permet une généralisation des balises utilisées (un catalogue de ces balises est calculé automatiquement à partir de la base).

- 95 Il allait par conséquent de soi de permettre à l'utilisateur de préserver une clause ou même un contrat dans sa base de clauses accessible en mode de rédaction. L'étape suivante consistait donc à permettre à chaque rédacteur de mettre ses propres créations à disposition de ses collègues du cabinet afin qu'ils les réutilisent le cas échéant.

2.5. Alertes structurelles

Un prototype d'aide complémentaire est en cours de développement. Il part du principe que la base de contrat est utilisée et que de nouveaux contrats ont été insérés. À partir d'outils simples utilisés dans le domaine de la linguistique de corpus, des calculs d'information mutuelle permettent de définir des cooccurrences fréquentes (collocations). Ici, nous les appliquons aux clauses au sein d'un type de contrat particulier dans un ordre donné ou pas, et à leur présence ou absence systématique dans certains types de contrats.

- 96 Ces calculs ont un double intérêt : d'une part, fournir au rédacteur des indications sur une forme de « bon usage » de la rédaction d'un type de contrat donné ; d'autre part, permettre d'améliorer les modèles structurels de contrats pour les schémas suggérés au rédacteur.
- 97 Par extension, ces mêmes outils pourront être utilisés dans la recherche automatique de motifs récurrents, selon différents critères, de manière à suggérer les formulations les plus fréquentes.

2.6. Intégration TAL

Par ailleurs, d'autres outils sont en cours de développement : tout d'abord un lexique de 460 000 formes fléchies. Chaque clause est traitée au fur et à mesure de son insertion dans le contrat. Si un mot n'est pas reconnu, il est délimité et les mots graphiquement les plus proches sont suggérés comme corrections.

- 98 Il est aussi nécessaire et judicieux d'intégrer à une telle plateforme un outil de correction grammaticale² – la base de contrats fournie pour le projet en est la preuve. Toutefois, les correcteurs existants, majoritairement commerciaux, sont adaptés à des écrits plus courants et moins marqués stylistiquement que l'écrit juridique. En la matière, les travaux de Souque (2008) ouvrent des perspectives pertinentes : une correction grammaticale fondée non pas sur des listes d'erreurs mais sur une recherche des

incohérences ; une base de règles ouverte et adaptable. Les listes d'erreurs ont en effet le défaut d'imposer un style et de détecter des erreurs là où le scripteur utilise des formulations désuètes ou stylisées. La possibilité de modifier les règles de détection d'incohérence grammaticale permet aussi d'envisager une adaptation de celles-ci aux pratiques d'écriture en langue du droit, en partant de l'hypothèse de l'existence d'une langue spécifique du droit, voire du contrat (Cornu, 2000).

3. Perspectives

Nous avons décrit ici SIRC, un système d'aide à la rédaction de contrats, se fondant sur une étude structurelle et linguistique de contrats et mettant à profit une base de données de contrat afin d'accompagner le rédacteur dans sa tâche. Toutefois, il s'agit d'une première implantation et de nombreuses perspectives sont envisageables, tout en apportant du grain à moudre aux domaines de la linguistique.

- 99 Un tel système est un moyen de rebondir sur des thématiques de recherches plus générales, tout en se donnant un cadre et un rattachement applicatif.
- 100 Au-delà du contrat de licence de propriété intellectuelle, les régularités observées et les réseaux référentiels peuvent-ils être observés au sein d'autres formes de contrats, voire au sein d'autres formes d'écrits juridiques ? Et dans quelle mesure la méthode d'aide à la rédaction développée pour SIRC peut-elle être étendue à d'autres écrits du droit ? Nous avons montré, non pas en ce qui concerne la génération du texte, mais du point de vue analyse automatique et extraction de termes, que les termes à portée juridique sont peu présents dans les textes législatifs (Lebarbé, 2007) car, une fois introduits, ils sont référencés par un système de pointage tout aussi complexe que celui constaté dans les contrats.
- 101 Ainsi il sera nécessaire de questionner la problématique du multilinguisme : comment intégrer les langues étrangères dans un système d'aide à la rédaction juridique ? Il semble difficile d'envisager l'appariement direct entre clauses traduites, bien que de nombreux systèmes de bases de traduction soient mis à la disposition des traducteurs. Non seulement il est fort probable que les clauses diffèrent dans leur style, mais aussi qu'elles varient en conséquence dans le fond. Cette tâche sera d'autant plus complexe qu'il faut différencier langue et pays. Les langues anglaise et américaine sont relativement similaires mais le système juridique diverge, notamment dans le cas de la propriété intellectuelle. Il faudra donc envisager un interfaçage entre clauses de langues différentes à l'aide de concepts pivots comme le propose Sérasset (2004) dans sa plateforme PAPILLON, mais d'un point de vue multilingue uniquement lexical. De même les réflexions, ne serait-ce qu'en terme de terminologie juridique multilingue (Gémar, 1995), montrent la difficulté de la tâche et l'extension du terme à la clause n'est pas chose acquise. Par ailleurs, l'impact de cette forme de normalisation multilingue entre clauses pourrait avoir des conséquences difficilement analysables *a priori* sur l'interprétation même des textes selon les juridictions.
- 102 Certains outils de linguistique de corpus ont été mentionnés. Leur utilisation est prévue pour aider les rédacteurs. Ces mêmes outils ont aussi pour objectif d'aider le linguiste à définir des prototypes sur plusieurs niveaux : prototypes structurels, stylistiques, syntaxiques et terminologiques. La base documentaire constitue une ressource en

perpétuelle expansion qui permettra d'affiner ces modèles, selon une typologie précise des contrats.

- 103 Enfin, la définition de ces modèles contribuera à une certaine standardisation terminologique et structurale du contrat ainsi qu'à une homogénéisation du discours juridique écrit. De manière intrinsèque au cabinet mais, plus largement au niveau international, nous percevons une harmonisation linguistique dans la lignée de certains projets nationaux ou européens, tels le projet LexAlp d'harmonisation terminologique en droit territorial à travers l'Arc Alpin, comme le montrent, dans ce numéro de *Lidil*, les articles de Lavault et Grossmann d'un point de vue linguistique et Randier et Voltmer d'un point de vue juridique.
- 104 Le domaine interdisciplinaire de l'informatique linguistique juridique, et notamment l'aide à la rédaction de documents juridiques, ne peut se concevoir sans l'interaction et le dialogue pluridisciplinaire du juriste, du linguiste et de l'informaticien. Une plateforme comme SIRC permet cette interaction en répondant aux besoins du juriste tout en apportant au linguiste la matière première pour étudier et modéliser la langue du droit, ce qui, en retour, peut bénéficier au juriste.

BIBLIOGRAPHIE

Références bibliographiques et webographiques

- CORNU G., (2000) : *Linguistique juridique*, Paris, édition Montchrestien, Domat Droit Privé.
- COTTIN S. (2005) : Le service public de diffusion du droit français sur l'Internet, *Acta Universitatis Sibiu* (Roumanie), n° 1-2/2005, <<http://www.servicedoc.info/Le-service-public-de-diffusion-du,1703>>.
- FRÉROT C., BOURIGAULT D., FABRE C., (2003) : Marier apprentissage endogène et ressources exogènes dans un analyseur syntaxiques de corpus. Le cas du rattachement verbal à distances de la préposition « de », *Revue TAL 44-3 : Évolutions en analyse syntaxique*, Éditions Hermes-Lavoisier, p. 167-186.
- GÉMAR J.-C. (1995) : *Traduire ou l'art d'interpréter, Tome 2, Application : traduire le texte juridique*, Sainte-Foy, Presse de l'Université du Québec.
- GLOBAL JUSTICE INFORMATION SHARING INITIATIVE : <<http://it.ojp.gov/jxdm>>.
- LEBARBÉ T. (2007) : LexTract : extraction semi-automatique de termes à portée juridique, *Actes des « Journées de Linguistique de Corpus »*.
- LEBARBÉ T., BLANCHARD A. et MEYNARD C. (2008) : Les manuscrits de Stendhal, de la base de données à la base documentaire, *Recherche et Travaux*, n° 72/2008, « De l'hypertexte au manuscrit », ELLUG, p. 97-117.
- LEUF B. et CUNNINGHAM W. (2001) : *The Wiki Way: Quick Collaboration on the Web*, Boston, Addison-Wesley.
- LEGALXML (2005) : OASIS not-for-profit global consortium, <<http://www.legalxml.org>>.

SÉRASSET G. (dir.)(2004) : *PAPILLON 2004: 5th Workshop on Multilingual Lexical Databases*, Grenoble, France.

SOUQUE A. (2008) : *Vers une nouvelle approche de la correction grammaticale, Actes de la Conférence TALN/RECITAL*, Avignon, 9-13 juin 2008.

NOTES

1. « Le contrat est synallagmatique ou bilatéral lorsque les contractants s'obligent réciproquement les uns envers les autres », Art. 1102, Code Civil.
 2. La distinction entre correction orthographique et correction grammaticale est, d'un point de vue informatique, à distinguer de la même distinction faite en didactique. Du point de vue informatique, la correction orthographique se limite à la validation de la présence ou pas d'un mot au sein d'un lexique.
-

RÉSUMÉS

La rédaction de documents juridiques obéit à un certain nombre de règles fondamentales allant de la structure globale du document aux choix syntaxiques et terminologiques. Le contrat en est un exemple, le contrat de licence ou de cession d'un droit de propriété intellectuelle est un cas particulier duquel peuvent être tirées des généralisations.

Nous présentons ici un travail applicatif interdisciplinaire qui prend en considération certains de ces paramètres d'ordre linguistique, et met leurs résultats à disposition du rédacteur de contrats. Par le biais d'une interface Web, non seulement le rédacteur, mais aussi ses interlocuteurs (les contractants), peuvent accéder à une aide à la rédaction dudit contrat. Celle-ci se fonde sur des prototypes rédactionnels, mais permet aussi un suivi de la rédaction du contrat et des alertes en cas d'incohérence structurelle.

Ce compte-rendu d'une collaboration entre université et entreprise montre avant tout qu'une informatisation d'observations d'ordre linguistique peut contribuer à un meilleur contrôle et une meilleure gestion de l'écrit juridique, au-delà de la sous-catégorie étudiée : le contrat de licence.

Legal document writing obeys to some fundamental rules ranging from global document structure to syntactic and terminology choices. Contracts are an example, licensing contracts or intellectual property rights assignment contracts are a particular case from which generalisations can be drawn.

We present here an applied interdisciplinary work that takes into account some linguistic parameters and makes them available to the contract writer. By means of a Web interface, the contract writer, but also his interlocutors (the contracting parties), can access to a writing assistance for the contract. The latter is based on redaction prototypes, but allows a collaborative approach of the contract writing and alarms in case of structural incoherence.

This account reporting a collaboration between university and industry shows above all the computational modelisations of linguistic observations can contribute to a better control and management of légal writings, much further than the studies sub-category: licence agreements.

AUTEUR

THOMAS LEBARBÉ

Université Stendhal–Grenoble 3, Laboratoire LIDILEM

BP15, 38040 Grenoble Cedex 9, France

thomas.lebarbe@u-grenoble3.fr