

Université de Montréal
et
Université Paris-Sorbonne

Résumé automatique de textes juridiques

par

Atefeh Farzindar

Thèse de doctorat effectuée en cotutelle

au

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences
Université de Montréal
et
École Doctorale Concepts et Langages
Université Paris IV-Sorbonne

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en informatique
et à

Université Paris IV-Sorbonne en vue de l'obtention du grade de Docteur
en informatique et sciences humaines

Décembre, 2004

©Atefeh Farzindar, 2004

Université de Montréal
Faculté des études supérieures
et
Université Paris-Sorbonne
École Doctorale Concepts et Langages

Cette thèse intitulée

Résumé automatique de textes juridiques

présentée et soutenue à
l'Université de Montréal et l'Université Paris-Sorbonne par

Atefeh Farzindar

JURY

Philippe Langlais (président), professeur à l'Université de Montréal
Guy Lapalme (directeur), professeur à l'Université de Montréal
Jean-Pierre Desclés (directeur), professeur à l'Université Paris-Sorbonne
Inderjeet Mani (rapporteur), professeur à Georgetown University
Amr Hilmy Ibrahim (rapporteur), professeur à l'Université de franche comté
Esma Aïmeur (examineur), professeur à l'Université de Montréal
Agatha Jackiewich (examineur), professeur à l'Université Paris-Sorbonne
Brahim Djioua (examineur), professeur à l'Université Paris-Sorbonne
**Ejan Mackaay, représentant du doyen de la Faculté des études supérieures
de l'Université de Montréal**

Thèse acceptée le : 29 mars 2005

Sommaire

Dans le cadre de notre thèse, nous avons travaillé sur la production automatique de résumé pour le domaine juridique. Nous avons développé le système LetSum (Legal Text Summarizer), un outil informatique en vue d'extraire l'information pertinente sous une forme condensée. Nous avons collaboré avec les avocats du centre de recherche en droit public (CRDP), chargés de créer une bibliothèque de droit virtuelle des décisions judiciaires canadiennes.

Notre méthode est basée sur l'analyse manuelle des jugements et de leurs résumés rédigés par les résumeurs professionnels. Dans cette méthode, nous identifions d'abord le plan d'organisation d'un jugement et ses différents thèmes discursifs qui regroupent les phrases autour d'un même sujet. Chaque phrase dans un thème donne des informations complémentaires sur le sujet. Pour les phrases reliées à un thème, nous pouvons interpréter les sens des phrases d'après leur contexte afin d'en extraire les idées clés.

La création du résumé par LetSum se fait en quatre étapes :

Segmentation thématique qui détermine l'organisation du document original et encadre les segments du texte associés avec un thème précis dans la jurisprudence.

Pour la structuration thématique, nous avons divisé un jugement en sept thèmes :

- **DONNÉES DE LA DÉCISION** : donne la référence complète de la décision et la relation entre les parties sur le plan juridique.
- **INTRODUCTION** : qui ? a fait quoi ? à qui ?
- **CONTEXTE** : recompose l'histoire du litige et l'histoire judiciaire.
- **SOUSSION** : présente le point de vue d'une partie sur le problème.
- **ISSUES**(ou les questions de droit) : identifient le problème juridique dont le tribunal est saisi.
- **RAISONNEMENT JURIDIQUE** : décrit l'analyse du juge, la détermination des faits et l'expression des motifs de la solution retenue.
- **CONCLUSION** : présente la décision finale de la cour.

Filtrage qui identifie les segments qui peuvent être supprimés dans les documents, sans perdre les informations pertinentes pour le résumé. Dans un jugement, les citations occupent un volume important du texte soit 30% du jugement, alors que leur contenu est moins important pour le résumé donc, nous identifions les citations principalement pour les supprimer.

Sélection des unités textuelles candidates pour le résumé qui construit une liste d'unités saillantes candidates pour chaque niveau structural du résumé en calculant les poids pour chaque phrase dans le jugement. La sélection est basée sur

des règles sémantiques et des calculs statistiques.

Production du résumé qui choisit les unités pour le résumé final et les combine afin de produire un résumé d'environ 10% du jugement.

L'évaluation de 120 résumés automatiques par 12 avocats du CRDP montre la qualité des résumés produits par LetSum qui est jugée excellente. Notre comparaison des résumés de LetSum avec cinq systèmes de recherches ou commerciaux met en évidence l'intérêt d'utiliser d'un système de résumé spécialisé pour le domaine juridique.

Mots-clés : résumé automatique, segmentation thématique, table de résumé, extraction des unités, texte juridique.

Abstract

In this thesis, we have focused on a problem referred to as automatic production of legal text summarization. We have developed a summarization system, called LetSum, for producing short summaries for legal decisions, which record the proceeding of a court. We have collaborated with the lawyers of the Public Law Research Center of Université de Montréal, developing a virtual library of Canadian law.

Our method is based on the manual analysis of the judgments by comparing manually written summaries and source documents to look for a match between the information considered important in the professional abstract and the information in the source decisions. Our approach investigates the extraction of the most important units based on the identification of thematic structure of the document and the determination of discursive themes of the textual units in the judgment. Each sentence in a theme gives additional information on the subject. For the sentences connected to a theme, we can interpret the meanings of the sentences according to their context to extract the key ideas.

The production of the summary is done in four steps :

Thematic segmentation detects the organizational architecture of a judgment. The document structure is based on the specific knowledge of the legal field. According to our analysis, we distinguish seven themes which divide the legal decisions into thematic segments :

- **DECISION DATA** : gives the complete reference of the decision and the relation between the parties for planning the decision.
- **INTRODUCTION** : who ? did what ? to whom ?
- **CONTEXT** : recomposes the story from the facts and events between the parties and findings of credibility on the disputed facts.
- **SUBMISSION** : presents the point of view the parties on the problem.
- **ISSUES** : identifies the questions of law for the court.
- **JURIDICAL ANALYSIS** : describes the comments of the judge, finding of facts and solution to the problem of the parties.
- **CONCLUSION** expresses the final decision of the court.

Filtering identifies parts of the text which can be eliminated, without losing relevant information for the summary. In a judgment, the citation units (sentence or paragraph) occupy a large volume in the text, up to 30%, of the judgment, whereas their contents are less important for the summary.

Selection builds a list of the best candidate units for each structural level of the

summary. Selection is based on the semantic rules and statistical computing.

Production of the summary chooses the units for the final summary and combines them in order to produce a summary of about 10% of the judgement.

The evaluations of 120 summaries by 12 lawyers show the quality of summaries produced by LetSum, which are judged excellent. Our comparison of the summaries of LetSum with five other research and commercial systems brings to the light the interest of using a system developed specifically for the summarization of legal documents.

Keywords : automatic text summarization, thematic segmentation, summary table, extraction of textual unity, legals texts

Table des matières

1	Introduction	1
1.1	Travaux préliminaires	3
1.2	Document Understanding Conference	3
1.2.1	DUC 2002	4
1.2.2	DUC 2003	5
1.2.3	DUC 2004	6
1.3	Résumé de la thèse	8
1.4	Résultats de nos travaux de recherche	9
1.4.1	Évaluation de LetSum	11
1.5	Plan de la thèse	12
2	Résumé automatique	14
2.1	Types de résumés	15
2.2	Méthodes de résumé automatique	16
2.2.1	Création automatique de résumé par extraction	17
2.2.2	Génération automatique de résumé	24
2.2.3	Résumé automatique de multi-documents	25
2.3	Discussion	26
3	Résumé de documents juridiques	28
3.1	Types de résumés juridiques	31
3.2	Ressources juridiques	32

3.3	Travaux en résumé de texte juridique	32
3.3.1	Production de résumé manuel	33
	Jurisprudence Express	33
	Fiches analytiques	33
3.3.2	Approches automatiques de production de résumé	36
3.4	Discussion des méthodes de génération de résumé	38
4	Étude de corpus juridique	39
4.1	Corpus d'étude	39
4.1.1	Composition	39
4.1.2	Structuration des textes juridiques	40
4.2	D'où viennent les informations	43
4.3	Dictionnaire de connaissances linguistiques	45
4.4	Conclusion	46
5	Méthode de constitution automatique de fiches de résumé	47
5.1	Introduction	47
5.2	Étapes de la constitution de la fiche de résumé	48
5.2.1	Analyse thématique	48
	Expérimentation	48
	Module de segmentation thématique	50
5.2.2	Filtrage	54
5.2.3	Sélection des unités textuelles	56
5.2.4	Production du résumé	58
5.3	Algorithme	58
5.4	Un exemple d'une fiche de résumé	60
6	Implémentation de LetSum	62
6.1	Introduction	62

6.2	Architecture de GATE	63
6.3	Implémentation	64
6.4	Prétraitement	65
6.5	Segmentation thématique	67
6.5.1	Identification du thème INTRODUCTION	69
6.5.2	Identification du thème ISSUE	70
6.5.3	Identification du thème CONTEXTE	71
6.5.4	Identification du thème SOUMISSION	72
6.5.5	Identification du thème CONCLUSION	73
6.5.6	Identification du thème RAISONNEMENT	74
6.6	Filtrage	75
6.6.1	Détection des citations	75
6.7	Sélection des unités saillantes	76
6.7.1	Module de calculs statistiques	81
6.8	Production de la fiche de résumé	82
6.9	Sorties de LetSum	83
6.10	Conclusion	87
7	Évaluation des systèmes de production de résumé	88
7.1	Introduction	88
7.1.1	Évaluation intrinsèque	89
7.1.2	Évaluation extrinsèque	90
7.2	Campagnes d'évaluations	91
7.2.1	Évaluation de SUMMAC	91
7.2.2	Évaluation de DUC	92
	Évaluation automatique avec ROUGE	92
7.3	Évaluation de LetSum	93
7.4	Évaluation des modules de LetSum	93

7.4.1	Évaluation de LetSum par ROUGE	95
7.5	Évaluation extrinsèque de LetSum	97
7.5.1	Évaluation manuelle	98
7.5.2	Évaluation par les juristes	100
7.6	Résumé	103
8	Perspectives futures	106
8.1	Génération de titres significatifs pour les jugements	106
8.2	Aspects temporels dans le résumé	108
8.3	Production de résumé pour les autres types de jugement	110
8.4	Traitement des jugements en français	110
9	Conclusion	111

Liste des tableaux

1.1	Douze questions qualitatives utilisées par NIST lors de l'évaluation de DUC	5
4.1	Alignement des unités textuelles du texte original du jugement avec les unités du résumé rédigé manuellement	42
4.2	Structure thématique d'une jurisprudence	44
4.3	Les pourcentages de la contribution de chaque thème dans les résumés et les jugements originaux	45
5.1	Les marqueurs significatifs qui se trouvent dans les titres des sections	51
5.2	Un exemple de segment de DONNÉES DE LA DÉCISION dans un jugement	52
5.3	Exemple d'une citation	55
5.4	Les marqueurs linguistiques de citation	56
5.5	Les marqueurs linguistiques utilisés dans le module de sélection des unités textuelles qui signalent les phrases importantes dans le jugement	57
5.6	Exemple de l'analyse d'un jugement pour construire une fiche de résumé	61
6.1	Exemple d'une phrase analysée par l'analyseur syntaxique	66
6.2	Identification et classification des titres significatifs par thèmes	67
6.3	Identification du corps du jugement	68
6.4	Règles pour la segmentation thématique : thème INTRODUCTION	69
6.5	Règle pour la segmentation thématique : thème ISSUE	70
6.6	Règles de segmentation thématique : thème CONTEXTE	71
6.7	Règles de segmentation thématique : thème SOUMISSION	72
6.8	Règles de segmentation thématique : thème CONCLUSION	73

6.9	Règle pour la segmentation thématique : thème RAISONNEMENT	74
6.10	Grammaire d'identification des unités de citation	76
6.11	Grammaire d'identification des phrases saillantes de INTRODUCTION	77
6.12	Grammaire d'identification des phrases saillantes de CONTEXTE	78
6.13	Grammaire d'identification des phrases saillantes de RAISONNEMENT	79
6.14	Grammaire d'identification des phrases saillantes de CONCLUSION	80
7.1	Résultat d'évaluation intrinsèque avec ROUGE-N et ROUGE-L	97
7.2	Questions juridiques et résultats d'évaluation extrinsèque	101
7.3	Résultats d'évaluation extrinsèque	103
8.1	Génération d'un titre informatif pour un jugement	108
9.1	Évaluation d'un juriste sur les résumés de la trousse d'évaluation	122

Table des figures

1.1	L'utilisation de l'information de <i>Background</i> pour garder la trace dans une série d'événements dans les articles journalistiques (Montréal Gazette 19/02/2003).	7
1.2	Fiche de résumé produit par LetSum, composé de 350 mots alors que le jugement source avait 4500 mots	11
2.1	Un exemple de règle écrite dans le langage formel	21
3.1	Exemple d'un résumé manuel écrit par SOQUIJ	34
3.2	Une fiche analytique d'une décision de la Cour fédérale	35
5.1	Les étapes de la constitution de la fiche de résumé	49
6.1	Sortie de module de segmentation thématique identifiant les thèmes du jugement, sous l'interface de GATE	83
6.2	Sortie de module de filtrage identifiant les citations, sous l'interface de GATE	84
6.3	Sortie de module de sélection identifiant les phrases importantes, sous l'interface de GATE	85
6.4	Fiche de résumé produit par LetSum, composé de 350 mots alors que le jugement source avait 4500 mots	86
7.1	Graphe de comparaison des résultats d'évaluation intrinsèque avec ROUGE	98
7.2	Graphe de comparaison des résultats d'évaluation extrinsèque sur les questions juridiques	102
7.3	Graphe de comparaison des résultats d'évaluation pour les valeurs qualitatives	104

7.4	Grphe de comparaison des r�sultats d'�valuation pour la valeur d'acceptabilit� sur la qualit� g�n�rale du r�sum�	105
-----	--	-----

Remerciements

Mes remerciements s'adressent tout d'abord à mon directeur de recherche à l'Université de Montréal, professeur Guy Lapalme pour avoir accepté de me diriger patiemment, pour son soutien constant pendant la rédaction de cette thèse, sa confiance et ses nombreux conseils.

Je remercie vivement mon directeur de recherche à l'Université Paris-Sorbonne, professeur Jean-Pierre Desclés qui m'a accueilli dans le laboratoire LaLICC dès le début de mon D.E.A., pour ses conseils judicieux et sa confiance en moi.

Ma gratitude au membre du jury, Philippe Langlais, Esma Aïmeur et Inderjeet Mani de m'avoir fait l'honneur d'évaluer ce travail.

J'ai trouvé au sein de l'équipe RALI un lieu propice aux échanges, aux discussions et aux collaborations scientifiques. Ce qui m'a offert la meilleure ambiance pour mon développement intellectuel. Un remerciement spécial à Elliott Macklovitch le responsable, Jian-Yun Nie et Philippe Langlais les professeurs du laboratoire RALI.

Je tiens à remercier l'équipe LexUM du laboratoire d'informatique juridique du Centre de recherche en droit public de la faculté de droit de l'Université de Montréal. Je remercie pour leur collaboration et leur participation lors de mes évaluations, en particulier Daniel Poulin professeur et responsable de l'équipe et Frédéric Pelletier avocat de LexUM pour leurs conseils permettant d'améliorer mon projet doctoral.

Je tiens à remercier également les institutions suivantes pour leur soutien financier lors de mon étude de doctorat : la Fondation canadienne de l'innovation et du ministère de l'Éducation du Québec, le Conseil de recherches en sciences naturelles et en génie du Canada, les Laboratoires Universitaires Bell et l'Université de Montréal.

J'adresse un remerciement plus spécial à mon mari Massoud et ma fille Tina pour leur confiance, leur amour inconditionnel et leur encouragement à persévérer.

Un grand merci enfin à ma famille qui m'a permis d'étudier dans un pays lointain et m'a encouragée à chaque instant.

*À mon copain, ami et mari Massoud
To my parents Akram and Hossein*

Chapitre 1

Introduction

De jour en jour, la masse d'information textuelle sous forme électronique ne cesse d'augmenter, que ce soit sous forme de documents accessibles sur internet, dans les bases de données des entreprises et des gouvernements. Il devient de plus en plus difficile d'accéder aux informations intéressantes sans l'aide d'outils spécifiques.

Dans ce contexte, il est nécessaire de pouvoir accéder au contenu des textes par des moyens rapides et efficaces. C'est la fonction d'un résumé qui constitue un moyen efficace et éprouvé pour représenter le contenu des textes, et ainsi permettre un accès rapide à leur contenu. La technologie du résumé automatique qui permet de choisir un ou plusieurs documents utilement est une réponse à ce besoin.

Le but d'un système de résumé automatique est de produire une représentation abrégée couvrant l'essentiel du contenu d'une source d'information. Dans un contexte d'opérationnel, il faut aussi considérer les besoins de l'utilisateur et de la tâche demandée.

On ne peut pas imaginer dans notre vie quotidienne, une journée sans résumé, comme le souligne Inderjeet Mani [Mani, 2001]. Les titres des journaux sont des résumés des différents événements. Le texte d'un article de journal contient aussi un résumé : le premier paragraphe décrit brièvement ce qui est nouveau dans l'histoire d'un événement. Les articles scientifiques sont accompagnés de leurs résumés écrits par leurs auteurs. Les bulletins d'informations, les prévisions météo, les tables des résultats des compétitions sportives, les catalogues des bibliothèques, les catalogues des rabais d'un magasin, tous sont des exemples de résumés.

L'utilisation de la technologie du résumé automatique réduit les temps de recherche pour trouver les documents pertinents et allège le traitement des longs textes en iden-

tifiant les informations clés. Certaines techniques de filtrage d'information peuvent produire des résumés flexibles en fonction des intérêts de l'utilisateur.

Un résumé cohérent, lisible et bien rédigé par un résumeur humain (un spécialiste effectuant des résumés de documents) est-il toujours préférable à résumé automatique ? Pour répondre à cette question, il faut considérer plusieurs paramètres dont le volume d'informations à traiter. Il faut souvent fouiller dans des centaines de documents pour trouver un document avec le contenu souhaité. Pour résumer un long texte, il faut pour un humain le temps de lire le texte et l'expertise nécessaire pour fournir les renseignements sur un sujet il faut **analyser**, **sélectionner** les segments et les **recomposer** en particulier pour construire un résumé. Il faut aussi tenir compte des limites des humains pour certaines tâches répétitives, la fatigue, etc. De plus un bon résumeur humain n'est pas toujours disponible. Tous ces facteurs augmentent le coût d'un résumé manuel.

Pour toutes ces raisons, l'ordinateur peut parfois donner des résultats plus intéressants qu'un humain. L'ordinateur peut être utilisé comme aide à l'organisation des connaissances des textes et comme un outil pour les tâches de traitement de textes assisté par ordinateur.

De nombreux travaux basés sur les techniques d'Intelligence Artificielle (IA) et de Traitement Automatique des Langues Naturelles (TALN), ont proposé des méthodes pour créer automatiquement des résumés. Plus récemment, la série de compétitions *Document Understanding Conference* [DUC, 2004] a été organisée pour évaluer les systèmes de résumé automatique. Mani [Mani, 2001] [Mani et Maybury, 1999], Marcu [Marcu, 2000] et Minel [Minel, 2002] présentent des introductions à la situation actuelle dans ce domaine. Ces travaux seront présentés plus en détail au chapitre 2.

La dernière rencontre annuelle sur le résumé automatique tenue lors de la conférence *ACL 2004 Text Summarization Branches Out*, avait pour thème les résumés dans les domaines spécialisés. Certaines applications exigent une analyse profonde afin de dégager les particularités des domaines et d'identifier les aspects importants relatifs aux applications. Il est de plus en plus difficile d'accepter un résumé général pour toutes sortes de textes sans prendre en compte les connaissances spécifiques du domaine et les besoins des utilisateurs. Lors de cette conférence, nous avons présenté le domaine juridique comme étant un domaine ayant un grand besoin de résumés mais avec des exigences spécifiques. Nous nous intéressons au traitement de documents de type juridique en analysant les décisions des cours judiciaires du Canada. Dans cette thèse, nous explorons la problématique du traitement des textes juridiques, nous présentons notre méthode de production automatique de résumé de jugements permettant aux

juristes de consulter rapidement les idées clés d’une décision juridique pour trouver les jurisprudences pertinentes à leurs besoins. Nos résultats montrent que cette méthode de production automatique de résumé juridique est efficace.

1.1 Travaux préliminaires

Avant de commencer ce projet de résumé juridique, nous avons travaillé sur deux projets de production de résumé. Le premier est le système ContextO qui traite les textes en français [Minel *et al.*, 2001]. Ces travaux ont été menés au sein de l’équipe de Jean-Pierre Desclés, LaLICC (Langages, Logiques, Informatique, Cognition et Communication) à l’Université Paris-Sorbonne, qui utilise la méthode d’exploration contextuelle [Desclés *et al.*, 1991]. Le deuxième est le système SumUM développé à l’Université de Montréal [Saggion, 2001], basé sur une méthode d’analyse sélective pour les textes en anglais, pour lequel nous avons développé deux extensions pour participer aux campagnes d’évaluation DUC 2002 et DUC 2003.

1.2 Document Understanding Conference

Document Understanding Conference (DUC) est une série d’évaluations organisée par le *National Institute of Standards and Technology* (NIST) dans le cadre du projet TIDES (*DARPA’s Translingual Information Detection, Extraction, and Summarization*), dans le but de faire évaluer la technologie des résumés automatique et pour permettre à des chercheurs de participer à des expériences à grande échelle. Dans le cadre de DUC, l’organisateur détermine les niveaux de performance de systèmes minimaux (*baseline*) pour chaque tâche dans les différents niveaux de compétition et fournit des données de référence (documents et résumés correspondants) pour l’entraînement des systèmes.

Chaque année plus d’une vingtaine d’équipes participe à cette compétition. NIST envoie le corpus de test aux participants; qui disposent d’un temps limité (environ 10 jours) pour soumettre les résumés produits automatiquement au NIST pour l’évaluation. Depuis les trois dernières années, les organisateurs de DUC ajoutent de nouvelles tâches et adaptent le processus d’évaluation ainsi que le baseline pour ces tâches. Nous décrivons les compétitions de DUC 2002, 2003 et 2004 auxquelles notre équipe (RALI) a participé.

1.2.1 DUC 2002

Pour DUC 2002 dans le cadre de l'ACL 2002, il y avait les trois tâches suivantes :

1. Résumé automatique d'un mono document.
2. Résumé automatique de multi-documents : étant donné un ensemble de documents sur un sujet unique, les participants doivent créer 4 résumés génériques de l'ensemble avec approximativement 50, 100, 200, et 400 mots.
3. Un ou deux projets pilotes avec évaluation extrinsèque.

NIST a produit 60 ensembles de référence d'une dizaine de documents. Ces ensembles proviennent des données de TREC utilisées dans les compétitions de questions et réponses lors de TREC-9. Chaque ensemble contient des documents courts définis selon quatre types de critères : un événement simple de catastrophe naturelle, un événement simple dans n'importe quel domaine, des événements distincts multiples du même type, et des informations biographiques sur un même individu. NIST a effectué deux types d'évaluation des résumés : une faite par des humains pour les résumés générés et une autre faite automatiquement pour les extraits. Pour évaluer la qualité des résumés, NIST a utilisé douze questions qualitatives portant sur des aspects tels la grammaticalité, la cohérence et l'organisation des résumés.

Lors du workshop DUC pour évaluation manuelle, NIST a utilisé une version de SEE [Lin, 2001], qui permet de visualiser l'alignement manuel des unités de deux textes pour comparer et calculer la similarité entre les résumés modèles et les résumés produits afin de mesurer la couverture (*coverage*) c'est-à-dire la fraction du résumé modèle exprimée dans le contenu du résumé produit par le système. Pour évaluer les aspects tels que la grammaticalité, la cohérence et l'organisation des résumés, les douze questions qualitatives sont utilisées. La table 1.1 montre les questions utilisées par NIST lors de l'évaluation de DUC.

Pour cette compétition, nous avons participé avec une version modifiée de SumUM [Farzindar *et al.*, 2002b], le système de génération de résumé automatique des textes, qui avait été développé pour le résumé automatique de longs documents scientifiques et techniques [Saggion et Lapalme, 2002]. Malgré le fait qu'on avait utilisé SumUM sur un domaine complètement différent, les résumés produits par SumUM ont été jugés excellents par rapport à ceux produits par d'autres développés expressément pour ce genre de données et les résultats sont parmi les meilleurs de tous les systèmes qui ont participé à DUC 2002.

1. About how many gross capitalization errors are there ?
2. About how many sentences have incorrect word order ?
3. About how many times does the subject fail to agree in number with the verb ?
4. About how many of the sentences are missing important components (e.g. the subject, main verb, direct object, modifier) - causing the sentence to be ungrammatical, unclear, or misleading ?
5. About many times are unrelated fragments joined into one sentence ?
6. About how many times are articles (a, an, the) missing or used incorrectly ?
7. About how many pronouns are there whose antecedents are incorrect, unclear, missing, or come only later ?
8. For about how many nouns is it impossible to determine clearly who or what they refer to ?
9. About how times should a noun or noun phrase have been replaced with a pronoun ?
10. About how many dangling conjunctions are there ("and", "however" ...)?
11. About many instances of unnecessarily repeated information are there ?
12. About how many sentences strike you as being in the wrong place because they indicate a strange time sequence, suggest a wrong cause-effect relationship, or just don't fit in topically with neighboring sentences ?

TAB. 1.1 – Douze questions qualitatives utilisées par NIST lors de l'évaluation de DUC

1.2.2 DUC 2003

Pour DUC 2003 dans le cadre de HLT-NAACL 2003, les quatre tâches suivantes étaient définies :

1. Résumé très court (~10 mots) comme titre informatif d'un ensemble d'articles journalistiques,
2. Résumé court (~100 mots) pour un ensemble de multi-documents,
3. Résumé court (~100 mots) selon un point de vue donné,
4. Résumé court (~100 mots) de cet ensemble en réponse à une question, utilisant les questions de TREC.

Nous avons participé à l'évaluation de DUC pour une deuxième fois à partir d'une version modifiée du système SumUM, et nous avons concentré nos efforts sur le développement des possibilités de multi-documents et également ajouté quelques fonctions additionnelles pour produire des résumés courts en réponse à une question [Farzindar et Lapalme, 2003]. Nous avons proposé une approche pour le résumé de multi-documents d'articles journalistiques en utilisant des informations contextuelles (*Background*) pour produire un seul résumé pour un ensemble de documents.

Cette méthode est le résultat de nos observations sur un corpus journalistique dans le but de présenter une série d'événements pour les lecteurs humains. Nous avons étudié différents types de journaux pour comprendre ce qui est considéré comme une bonne méthode pour présenter une collection d'informations au sujet d'un concept spécifique à un lecteur et comment le lecteur peut combiner la nouvelle information de l'article courant avec l'information précédente pour parvenir à suivre une série d'événements.

Une bonne manière de suivre une succession d'événements dans une histoire est de maintenir certaines informations de contexte (background) à propos de ce sujet et de rechercher de nouvelles informations sur ces topiques dans un nouvel article. La figure 1.1 montre un exemple pris de la première page du journal de la Gazette de Montréal où peut être trouvé *Focus* de l'événement, le **Background**, **New** et **Next** avec le numéro de référence de page pour trouver plus de détails au sujet de cet événement. Cette organisation facilite la poursuite et la compréhension d'une série de nouvelles. Avec un background court au sujet de chaque événement, un lecteur peut facilement retrouver l'information nécessaire dans les documents reliés. Avec le temps, un lecteur obtient plus d'informations sur le contexte pour un nouveau document sur un concept précis qui vont l'aider pendant la recherche de nouvelles informations tout en évitant une certaine redondance qui apparaît nécessairement dans une série d'événements. Les résultats d'évaluation montrent une augmentation sur la précision et le rappel par rapport au DUC 2002.

1.2.3 DUC 2004

Pour le DUC 2004, les tâches de DUC 2003 ont été répétées et deux nouvelles tâches étaient ajoutées. La première nouvelle tâche était la production des résumés ciblés pour traiter la question de *Qui est X* afin de regrouper les informations essentielles sur un personnage.

La deuxième tâche additionnelle de DUC 2004 était le résumé translinguistique. La tâche consistait à produire un résumé très court en anglais d'articles journalistiques en arabe. Cette tâche traitait des phrases sorties d'un traducteur automatique de l'arabe vers l'anglais. Notre équipe RALI a participé à cette tâche avec une approche différente : création de résumé des documents arabes directement et traduction automatique du résumé arabe en anglais [Douzidia Soufiane et Lapalme, 2004]. Le résultat de l'évaluation a montré que cette méthode fonctionne mieux que celle des autres participants.



FIG. 1.1 – L’utilisation de l’information de *Background* pour garder la trace dans une série d’événements dans les articles journalistiques (Montréal Gazette 19/02/2003).

Les évaluations de DUC avant 2004 étaient faites manuellement. Pour la première fois lors de DUC 2004, l’évaluation automatique des résumés a été testée, en utilisant un logiciel développé à ISI par Chin-Yew Lin [Lin, 2004], implantant une mesure ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*). ROUGE est basé sur le calcul de similarité entre des résumés modèles, écrits par des humains, et des résumés automatiques. Nous décrivons cette évaluation en détail dans le chapitre 7.

En conclusion, il faut noter que les organisateurs de DUC ont commencé à encourager les recherches sur d’autres applications du résumé que celle consistant à produire des résumés d’articles journalistiques, car ces approches sont très limitées pour traiter les documents dans les vraies applications. Pour l’avenir de DUC (2005 jusqu’à 2007) un guide est soumis proposant d’élargir le champ de résumé pour de nouvelles applications.

1.3 Résumé de la thèse

La plupart des méthodes de résumé sont basées sur les articles journalistiques. Le domaine juridique est une application intéressante pour le résumé automatique, parce que les juristes consultent quotidiennement des milliers de documents juridiques stockés dans des bases de données volumineuses. À l'Université de Montréal, l'équipe LexUM, le laboratoire d'informatique juridique, est chargée de la conception et la gestion d'un système de publication gratuite sur internet des arrêts de jugements des cours canadiennes. Ils nous ont suggéré de travailler sur la collection de CanLII (Canadian Legal Information Institute), la bibliothèque de droit virtuelle, qui diffuse les décisions de tous les tribunaux judiciaires canadiens. Cette grande collection reçoit entre 10 000 et 15 000 visites par jour. Pour traiter un cas, les juristes cherchent les décisions des tribunaux sur lesquelles s'appuyer pour obtenir un traitement équivalent, ce qu'on appelle les **précédents juridiques**. L'ensemble des précédents juridiques forme la **jurisprudence**. Pour une seule affaire, il faut souvent étudier des dizaines de longs jugements. Sans l'aide d'outils informatiques, l'analyse de ces jugements demande beaucoup de temps et d'expertise.

Nous avons effectué une recherche afin de trouver les recherches en informatique juridique pouvant nous servir à compléter nos connaissances pour la création des résumés automatiques. Lorsqu'on analyse les caractéristiques des systèmes tels que décrits dans la littérature, on constate que la plupart des techniques calculent un poids pour chaque phrase individuellement et ils sélectionnent les phrases les plus importantes dans le texte. Ils ne tiennent toutefois pas compte des relations entre les phrases. Les phrases ne tiennent pas compte du contexte, ce qui peut entraîner de mauvaises interprétations comme nous le démontrons au chapitre 3.

Dans notre méthode, nous identifions d'abord le plan d'organisation d'un jugement et ses différents thèmes discursifs qui regroupent les phrases autour d'un même sujet. Chaque phrase dans un thème donne des informations complémentaires sur le sujet. Pour les phrases reliées à un thème, nous pouvons interpréter les sens sémantiques des phrases d'après leur contexte. Nous montrons qu'il y a certains thèmes qui contiennent des informations importantes dans un document. Nous identifions également les segments volumineux à supprimer dans le jugement qui n'apportent aucune information pertinente pour le résumé (voir chapitre 4).

1.4 Résultats de nos travaux de recherche

Notre approche est basée sur l'exploitation de la structure thématique d'un jugement. L'identification de l'organisation structurelle du document nous donne la possibilité de catégoriser les informations et de trouver les bornes de changement thématique dans le document afin d'en extraire les idées clés. Nous montrerons que la jurisprudence possède une structure thématique. Les différents thèmes dans le jugement découpent la suite des paragraphes en segments thématiques. Pour chacun, nous lui assignons un rôle rhétorique qu'il joue dans le jugement. Dans cette thèse, nous utilisons la notion de **thème** comme équivalent au rôle rhétorique pour chaque segment thématique. Cette étude montre l'intérêt de l'analyse des phrases dans leur contexte, ce qui permet l'extraction des phrases les plus importantes d'après leurs thèmes [Farzindar et Lapalme, 2004a].

Nous avons divisé un jugement en sept thèmes (voir chapitre 5) :

- **DONNÉES DE LA DÉCISION** présentent la référence complète de la décision : nom de la cour de décision, lieu de l'audience, date du jugement, numéro de greffe, référence neutre, identification des parties, intitulé du jugement, autorités et doctrines citées.
- **INTRODUCTION** qui ? A fait quoi ? À qui ?
- **CONTEXTE** recompose l'histoire du litige et l'histoire judiciaire.
- **SOUMISSION** présentent le point de vue d'une partie sur le problème.
- **ISSUES** (ou les questions de droit) identifient le problème juridique dont le tribunal est saisi.
- **RAISONNEMENT JURIDIQUE** contient l'analyse du juge, la détermination des faits et l'expression des motifs de la solution retenue.
- **CONCLUSION** est la décision finale de la cour.

Nous proposons également une nouvelle forme de présentation de résumé à l'utilisateur sous une forme d'une fiche de résumé, ce qui sert à l'organisation des informations extraites dans un format plus lisible et plus facile à utiliser. Cette fiche permet de présenter les informations considérées importantes associées aux thèmes précis, ce qui en facilite la lecture et la navigation entre le résumé et le jugement source. Pour chaque phrase du résumé produit, l'utilisateur peut en déterminer le thème ce qui aide à comprendre l'ordre des événements et le raisonnement du juge. Si une phrase semble plus importante pour l'utilisateur et qu'il désire plus d'information sur ce sujet, on peut lui proposer le segment thématique entier contenant la phrase sélectionnée, pour obtenir les informations complémentaires sur le sujet.

La constitution de la fiche de résumé se fait en quatre étapes : segmentation thématique, filtrage des unités moins importantes comme les citations des articles des lois, sélection des unités textuelles candidates pour le résumé et production du résumé selon la taille demandée.

Segmentation thématique : cette étape détermine l'organisation du document original et encadre les segments du texte associés avec un thème précis dans la jurisprudence. Notre segmenteur thématique est fondé sur les connaissances sémantiques du jugement. Afin de séparer un segment thématique et de détecter son thème, nous avons utilisé des conditions telles que la présence des titres des sections, les positions d'un segment, l'identification des styles direct et indirect et la présence des marqueurs linguistiques.

Filtrage : cette étape identifie les segments qui peuvent être supprimés dans les documents, sans perdre les informations pertinentes pour le résumé. Dans un jugement, les citations occupent un volume important dans le texte soit 30% du jugement, alors que leur contenu est moins important pour le résumé, donc nous les considérons comme des exemples négatifs. Pour cette raison, à l'intérieur des segments thématiques nous identifions les citations à supprimer. Nous supprimons aussi les soumissions des parties et leurs points de vue sur les événements qui sont loin du sujet de l'analyse directe du juge sur le litige.

Sélection des unités saillantes : cette étape construit une liste d'unités saillantes candidates pour chaque niveau structural du résumé. Il calcule un poids pour chaque phrase dans le jugement d'après les fonctions heuristiques basées sur les informations suivantes : la position des paragraphes dans le document, la position des phrases dans le paragraphe, les marqueurs linguistiques, les *cue-phrases*, les vocabulaires contrôlés du domaine juridique et les fréquences des mots dans le texte et le corpus.

Production de la fiche de résumé : Une fois sélectionnées les unités candidates potentielles pour le résumé, cette étape choisit les unités pour le résumé final et les combine afin de produire un résumé d'environ 10% du jugement. Le critère de sélection des unités est basé sur la pondération du segment thématique contenant les unités candidates. Selon nos analyses de corpus, nous avons constaté que la distribution de l'information dans les résumés des arrêtistes donne la possibilité de mesurer l'importance des segments thématiques. Lors de cette étape de sélection de la liste des unités candidates, nous choisissons les unités du segment thématique INTRODUCTION avec les scores plus élevés jusqu'à concurrence de 10% de la taille de résumé. Dans le segment CONTEXTE, les unités sélectionnées occupent 24% de la longueur du résumé. La contribution du segment RAISONNEMENT JURIDIQUE est 60% et les unités avec le

rôle CONCLUSION occupent 6% du résumé.

La figure 1.2 montre un exemple de sortie de LetSum comme une fiche de résumé. Cette fiche de résumé montre les différents thèmes identifiés dans le jugement et qui étaient pertinents pour le résumé. Le texte de résumé est de 10% de la taille de jugement original (le document source a dix pages).

RCMPT-979-96.html	
INTRODUCTION	[1] This is an application by Her Majesty the Queen (Crown) for an order striking out the Statement of Claim or, in the alternative, an extension of time to allow the Crown to file a Statement of Defence in the present action. [7] I believe, that before I recite the facts of the present case, it is important to note that on a motion to strike a Statement of Claim due to the fact that the Statement of Claim discloses no reasonable cause of action, it must be plain and obvious that the claim will not succeed notwithstanding the fact that the allegations in the Statement of Claim must be deemed to be true.
CONTEXT	[11] The plaintiff (Riabko) was a member of the Royal Canadian Mounted Police (RCMP) from November 6, 1978 to September 14, 1994, almost 16 years. On May 6, 1994 an Adjudication Board created under sections 43 and 44 of the According to the Crown "These actions arose from certain incidents in which the plaintiff was involved in and occurred in 1992". [13] As a result of the Board's decision of May 6, 1994, Riabko was sanctioned by requesting or ordering his resignation from the RCMP Force within 14 days. [16] On April 30, 1996, Riabko filed a Statement of Claim in this action in the Federal Court of Canada.
ISSUE	Issue[27] Does the Statement of Claim show a triable issue?
REASONING	I take this to mean that if the sections of the Act and Regulations are followed, a member may be dismissed or discharged and that the member would not be able to pursue the issue in the Courts by means of filing a Statement of Claim only alleging wrongful dismissal. [35] Because of the alleged breach of the RCMP Code of Conduct, a formal disciplinary hearing took place pursuant to section 43 of the RCMP Act, that is, an Adjudication Board was appointed to conduct a hearing into the alleged complaint. [42] It is obvious that the plaintiff Riabko did not follow the procedure set out in the RCMP Act and he is now alleging that he is claiming against Her Majesty because the process wherein he was asked to resign was an abuse of power by the Board, that is, from the very start, the process of the Board was flawed and he would thus have the right to proceed in Court. [45] I am satisfied that by having resigned, she could not avail herself of the internal process as stated in the RCMP Act and could sue for damages for sexual harassment. It must be noted that before she commenced her action before the Federal Court she did not avail herself or never took part in the process set out in the "She never did anything wrong" while in the case at bar the plaintiff was found to have contravened the RCMP Code of Conduct. [47] I am satisfied that where it cannot be shown that the power with regard to the grievance process as set out in the RCMP Act has been exceeded or abused, then there would be no cause of action. [49] I am satisfied there would be no purpose for Parliament to set out a grievance procedure by statute if a party could, after taking part in the procedure, decide to circumvent the statutory procedure.
CONCLUSION	[50] As well, after a plain reading of the Statement of Claim, and particularly paragraphs 5 and 6, I am satisfied that there is no allegation that the Adjudication Board of the RCMP abused or exceeded its jurisdiction. [51] Plaintiff's claim is struck with costs.

FIG. 1.2 – Fiche de résumé produit par LetSum, composé de 350 mots alors que le jugement source avait 4500 mots

1.4.1 Évaluation de LetSum

Nous avons évalué LetSum en deux étapes : d'abord nous avons mesuré la performance des deux modules principaux de la segmentation thématique et du filtrage. Après nous être assuré des fonctionnalités de ces modules, nous avons évalué les modules de sélection des phrases et de production dans le cadre d'une évaluation globale des phrases extraites du résumé.

Nous avons évalué notre système de résumé par des évaluations intrinsèque et extrinsèque. Nous avons comparé différentes méthodologies de productions de résumé. Pour l'évaluation intrinsèque, nous avons utilisé la méthode d'évaluation automatique ROUGE [Lin, 2004] (voir chapitre 7) en utilisant les résumés manuels comme les résumés modèles. Nous avons compté le chevauchement de n-grammes et la plus longue sous-séquence commune entre les résumés automatiques et le résumé modèle.

Pour l'évaluation extrinsèque, nous avons profité de l'avis d'avocats du CRDP comme experts du domaine juridique et utilisateurs particuliers du système en mesurant les informations pertinentes exprimées dans les résumés à l'aide de réponses à des questions sur les points juridiques. Notre hypothèse est que si un expert juridique est capable de répondre à une série de questions sur les idées principales du jugement en lisant le résumé alors le résumé est assez complet et il contient des informations clés du jugement.

Nous avons aussi évalué les aspects qualitatifs du résumé comme la lisibilité, la cohérence, la pertinence des phrases extraites et l'acceptabilité du résumé. Les avocats ont évalué les résumés sans savoir lesquels avaient été rédigés par un humain et lesquels avaient été produits par des machines et par quelles méthodes. Dans toutes les évaluations, LetSum est jugé plus performant que les autres méthodes (expliquées au chapitre 7 section 7.4.1). Selon les avocats, les résumés produits par LetSum sont équivalents aux résumés rédigés par les arrêtières des cours.

L'évaluation manuelle basée sur les jugements humains est une évaluation précise sur les mesures de qualité de l'efficacité d'un système informatique comme un système de résumé automatique, mais elle est très coûteuse en matière de temps et de ressources humaines. Alors que les évaluations automatiques sont beaucoup moins chères. D'après nos expériences d'évaluation, nous avons observé que l'évaluation automatique avec ROUGE au niveau de comparaison des différents systèmes est très efficace et le résultat obtenu par ROUGE pour ordonner les systèmes d'après leurs performances est la même que celle que nous avons obtenue avec l'évaluation avec des juristes (voir chapitre 7).

1.5 Plan de la thèse

La thèse est organisée de la manière suivante. Dans le chapitre 1, nous avons présenté les résultats de l'expérience de la participation à la campagne d'évaluation DUC 2002 [Farzindar *et al.*, 2002b][Farzindar *et al.*, 2002a] et DUC 2003 [Farzindar et Lapalme, 2003] lors des workshop de conférences ACL 2002 (*Association for Com-*

putational Linguistics) et NAACL 2003 (*North American chapter of the Association for Computational Linguistics*). Dans le chapitre 2, nous introduisons d’abord la terminologie nécessaire, l’état de l’art sur la condensation automatique de texte et la problématique ainsi que les différentes méthodes de création automatique de résumé et évaluation de systèmes. Le chapitre 3 est consacré aux caractéristiques du domaine et du corpus ainsi que les travaux pertinents sur le résumé de texte juridique. Nous décrirons différentes méthodes de production de résumé juridique utilisées actuellement par des experts et les approches automatiques du point de vue de leur adéquation par rapport aux textes légaux. Le chapitre 4 détaille l’analyse du corpus et l’organisation des connaissances linguistiques pour la tâche de production des résumés. Ces travaux ont été présentés lors de la conférence TALN-RECITAL 2004 (Le Traitement Automatique des Langues Naturelles) [Farzindar, 2004]. Le chapitre 5 décrit notre méthode pour produire des résumés juridiques [Farzindar et Lapalme, 2004d], présenté à CLiNE 2004 (*Computational Linguistics in the North-East*). Notre article [Farzindar *et al.*, 2004] est publié dans la revue internationale Traitement Automatique des Langues (TAL), cet automne (2004) pour la première fois, cette revue publie un numéro spécial sur le résumé automatique pour en réunir quatre articles proposant des solutions et les perspectives sur ce problème. Le chapitre 6 présente notre prototype appelé LetSum (*Legal text summarizer*). L’article [Farzindar et Lapalme, 2004a] est publié lors de la conférence annuelle sur le résumé automatique (dans le cadre de la conférence ACL 2004) dont le thème était : les applications possibles de résumé dans les contextes de productions. Dans cet article, nous avons proposé un nouveau domaine pour le résumé avec un grand besoin des usagers et notre solution pour ce problème. Le chapitre 7 explique le problème de l’évaluation de résumés automatiques ainsi que les méthodes utilisées pour évaluer LetSum. [Farzindar et Lapalme, 2004c] a été présenté à la conférence Jurix, la plus grande conférence sur l’intelligence artificielle et le droit. Dans cet article, nous avons présenté notre système, LetSum et le résultat de l’évaluation manuelle par les avocats sur 120 résumés de jugements. Cette évaluation montre la performance de notre système parce que la qualité des résumés produits est très proche de celle des résumés écrits par un humain. L’article [Farzindar et Lapalme, 2004b] publié à CaSTA 2004 (*Canadian Symposium on Text Analysis*), présente LetSum au point de vue de sa fonctionnalité dans les sciences humaines et par rapport à des utilisateurs comme nous le découvrons au chapitre 7. [Farzindar et Lapalme, 2005], accepté à la conférences TALN 2005, est consacré à l’évaluation de qualité et d’acceptabilité des résumés automatiques. Un panorama des travaux futurs est présenté au chapitre 8. Finalement, le chapitre 9 présente des conclusions.

Chapitre 2

Résumé automatique

Le but d'un système de résumé automatique est de produire une représentation condensée d'une source d'information, où les informations importantes du contenu original sont préservées. Une source d'information peut désigner différents genres : document multimédia, e-mail, hypertexte, etc., utilisé dans divers domaines : scientifique, littérature, juridique, etc. Le résumé produit par le système peut se présenter sous différentes formes et types. Le résumé peut être produit à partir d'un seul document ou à partir de plusieurs documents reliés sur un même sujet. Les méthodes de création de résumé de texte automatique sont classées en deux approches : extraction et génération (abstraction). Le résumé produit par génération est le résultat d'un processus de compréhension du texte source qui exige des synthèses sur les grammaires et les lexiques pour l'analyse syntaxique et la génération, tandis que l'extraction est un processus de sélection des unités saillantes (mots, phrases, paragraphes, etc.) à partir de l'entrée originale [Mani *et al.*, 2002].

Un autre critère distinguant les systèmes est le degré de généralité du contenu du résumé : les résumés qui traitent tous les sujets d'un document avec le même degré d'importance sont des résumés génériques, tandis que des résumés orientés par une requête se concentrent sur les facettes spécifiques du document source qui sont reliées au sujet spécifié par la requête [Farzindar et Lapalme, 2003].

Nos études porteront sur une forme particulière de documents de type juridique, les décisions des cours judiciaires du Canada et les différents aspects de ce domaine d'application. L'objectif de ce projet est de développer un système de création automatique de résumés courts, qui répond aux besoins des avocats, des juges et des experts du domaine.

La majorité des recherches sur le résumé automatique est limitée au traitement des articles de presse. L'idée de création d'un résumé orienté par la tâche ou d'un résumé pour des utilisateurs spécifiques d'un domaine est plus récente, en particulier au sein des travaux futurs de la campagne DUC ¹. En 2004, la rencontre annuelle sur le résumé automatique (*Text summarization branches out*) avait pour but le thème de montrer l'utilité du résumé automatique pour des applications telles que le domaine juridique, le diagnostic médical ou les secours d'urgence et les besoins de santé publique dans les désastres naturels. Par exemple le chef d'un programme d'organisation des secours d'urgence cherche des informations spécifiques dans les textes décrivant les besoins dans le cas d'un tremblement de terre. Notre travail est le premier système complet de production de résumés de textes juridiques.

Dans ce chapitre, nous présentons l'état actuel des recherches et un aperçu des systèmes de résumé de texte automatique. Nous débutons par des définitions de base et des principes généraux du résumé. Par la suite, nous décrivons les méthodes d'extraction et de génération, ainsi que les approches de résumé de mono document et multi-documents. Finalement, nous situons notre recherche par rapport aux travaux actuels et futurs de la communauté de DUC.

2.1 Types de résumés

Des nombreux types de résumé ont été identifiés [Borko et Bernier, 1975] [Cremmins, 1996] [Spark Jones, 1999] [Hovy et Lin, 1999] selon leur longueur, leur style et leur subjectivité :

résumé indicatif qui fournit une idée du texte sans en donner le contenu spécifique, il signale les thèmes du document et son style peut être télégraphique.

résumé informatif qui renseigne sur les informations essentielles contenues dans le document. Les informations sont présentées dans l'ordre du document, mais leur importance relative peut différer de celle du document.

résumé d'auteur qui donne une présentation brève du contenu d'un article par son auteur.

résumé-synthèse qui présente les sujets principaux d'un long document.

résumé sélectif qui néglige les aspects très généraux d'un document et développe des parties spécialisées.

¹*Roadmap 2005-2007* <http://www-nlpir.nist.gov/projects/duc/>

sommaire qui énumère les points principaux d'un discours.

digest ou fiche analytique qui condense un livre ou un article de journal.

résumé ciblé qui se concentre sur les contres d'intérêt pour l'utilisateur, alors que le résumé générique reflète plutôt le point de vue d'auteur.

2.2 Méthodes de résumé automatique

L'idée de produire des résumés automatiques n'est pas nouvelle [Luhn, 1959]. Les approches classiques de résumé automatique incluent l'utilisation des indicateurs de surfaces de l'information pertinente et les statistiques de corpus qui peuvent être appliquées au texte sans restriction. Les premiers travaux [Luhn, 1959] [Edmundson, 1969] ont étudié les techniques suivantes :

Position dans le texte où les phrases sont pondérées selon leur position (premier paragraphe, ou *Introduction*, *Conclusion* ou la partie après les titres de section, etc...).

Sélections lexicales où la présence de marqueurs linguistiques est identifiée.

Lieu où les premières (et dernières) phrases de chaque paragraphe devraient contenir l'information sur le sujet.

Fréquence où les mots significatifs sont fréquents dans le document alors qu'ils ont une fréquence faible en général.

Bien que chacune de ces approches ait une certaine utilité, elles dépendent du format et du modèle d'écriture. La stratégie qui consiste à prendre le premier paragraphe n'est pas universelle, elle ne fonctionne que pour les articles journalistiques. Dans d'autres cas comme pour la production d'un résumé biographique ou d'une décision juridique, la position des phrases n'est pas toujours un élément prépondérant. Pour élaborer des systèmes de résumé automatique, la plupart des chercheurs se sont penchés sur deux types d'approches : par extraction ou par génération. Les extraits sont créés par réutilisation directe des parties (mots, phrases, etc) du texte d'entrée, alors que les résumés générés sont créés par régénération à partir d'une représentation abstraite du contenu extrait.

La production du résumé comporte habituellement les processus suivants :

extraction : identification des éléments importants dans le texte,

génération : reformulation du matériel extrait dans un nouveau texte cohérent,

fusion : combinaison des parties extraites,

compression : élimination des éléments de moindre importance.

2.2.1 Création automatique de résumé par extraction

Pour créer un résumé par extraction, il faut d'abord repérer les unités textuelles (expressions, propositions, phrases, paragraphes) considérées comme saillantes ; ensuite le système va sélectionner les unités textuelles qui portent les idées principales du contenu du document. Le résumé produit par extraction respecte l'ordre d'apparition des phrases dans le document original.

Les approches d'extraction récentes emploient des techniques plus sophistiquées pour extraire les phrases ; ces techniques se fondent souvent sur l'apprentissage machine pour identifier les indicateurs importants, sur l'analyse de langage naturel pour identifier les passages pertinents ou sur des relations entre les mots.

Nous allons présenter les approches importantes pour l'extraction des unités textuelles pertinentes.

Position des phrases

À cause de la structure des textes, certaines positions dans le texte risquent de contenir les informations importantes, donc la position relative d'une phrase dans le document est un bon indicateur de l'importance de la phrase. [Baxendale, 1958] a prouvé que les premières et dernières phrases des paragraphes sont jugées importantes : dans un corpus de 200 paragraphes, on a constaté que, dans 85% des paragraphes, la première position contient la phrase la plus importante au paragraphe et dans 7% des paragraphes la phrase la plus importante s'est produite en dernière position. Récemment, [Lin et Hovy, 1997] a développé des algorithmes pour l'identification automatique de la position afin de choisir de bonnes phrases pour le résumé pour créer une liste qui indique les positions dans le texte des phrases de contenant les topiques. La position physique de la phrase mais également sa position logique comme membre d'une section conceptuelle particulière est importante pour l'extraction de phrase. [Saggion et Lapalme, 2000] a montré que, dans un corpus des résumés écrits par les abstracteurs professionnels, plus de 70% d'information pour le résumé viennent de l'introduction, la conclusion, les titres principaux et les titres de section du document.

Expressions prototypiques

L'approche de [Edmundson, 1969] et [Paice, 1981] utilise des expressions courantes ou *cue phrases* (*cet article propose ... , en conclusion, ...*) qui signalent l'importance

de phrases qui les contiennent. Pour extraire ces phrases, il pondère ces expressions par leur position, leur fréquence relative dans le corpus, etc. Les *cue phrases* sont regroupées en classes positives et négatives. Dans le cas d'articles scientifiques, [Teufel et Moens, 1997] rapporte un taux de 54% (rappel et précision) avec l'utilisation d'une liste de 1 423 *cue phrases* identifiées manuellement.

Mots des titres

Edmundson [Edmundson, 1969] utilise cette hypothèse : si les principaux thèmes sont véhiculés dans les titres. Les phrases importantes sont donc celles qui contiennent les mots des titres et les scores attribués aux phrases sont fonction de l'occurrence de mots dans les titres et de leur regroupement.

Fréquence des mots

Luhn [Luhn, 1959] utilise la loi de Zipf pour la distribution des mots (quelques mots se produisent très souvent, peu de mots se produisent quelques fois, et beaucoup de mots se produisent rarement) pour développer ce critère : si un texte contient quelques mots peu fréquents, alors les phrases contenant ces mots sont probablement importantes. L'algorithme de Luhn filtre des mots dans le document en utilisant une liste des mots vides. Ensuite, les mots sont normalisés par similitude orthographique : une paire de mots (ex. similar et similarity) est assorti à chaque position de caractère. Une fois une disparité se produit, le nombre de lettres suivant le caractère non semblable de toutes les deux est compté. Si ce nombre est moins de 6, ces mots sont regroupés ensemble. Les fréquences de ces mots dans le document sont notées et les mots de basse fréquence sont enlevés. Les phrases sont pondérées en utilisant l'ensemble résultant de termes significatifs et d'une mesure de densité des mots. Chaque phrase est divisée en segments encadrés par des termes significatifs. Chaque segment est pondéré en prenant le carré du nombre de termes significatifs encadrés divisés par le nombre de termes significatifs. Le segment possédant le score le plus élevé est pris comme poids de la phrase. Ce qui est connu sous le nom de fréquence inverse de document (*idf*) a été défini par [Spark-Jones, 1972], initialement nommé le schéma du poids de terme (*term weighting scheme*). En traitant une collection des textes dans un certain domaine (par exemple : informatique) il est probable que tous les documents partagent des termes communs dans ce domaine (par exemple : ordinateur, algorithm) alors qu'ils sont moins fréquents dans la collection entière (par exemple : système distribué). Dans ce cas, la pertinence d'un terme dans le document est aussi reliée à l'inverse du nombre de documents dans

la collection contenant le terme [Salton, 1988]. La formule normalisée pour la pertinence de terme est $tf * idf$, pour chaque mot du texte à résumer, une fois éliminer les mots vides. La technique de pondération utilisée pour calculer du poids d'une phrase est :

$$score(M) = f_{local} * \log(100 * N / f_{global})$$

f_{local} : fréquence du mot dans le texte ;

f_{global} : fréquence du mot dans le corpus de référence ;

N : nombre de documents du corpus ;

Par la suite, un score est attribué à chaque unité textuelle (ex : phrase) par addition des scores de chacun des mots contenus dans celle-ci. Comme alternative à la pertinence de terme, la pertinence de concept peut mesurer en utilisant WordNet [Lin et Hovy, 1997] (une occurrence du concept *fruit* est comptée quand le mot *pomme* est trouvé comme un hyperonyme de *fruit*).

Connexion par chaînes lexicales

Cette approche exploite des techniques de TAL, ainsi que des relations sémantiques entre les différents mots en se basant sur la notion de concept [Hahn, 1990]. La méthode de [Barzilay et Elhadad, 1997] identifie des *chaînes lexicales* entre les mots et sélectionne les phrases par chaînes lexicales. Les mots peuvent être connectés par différents types de relations sémantiques : synonymes, hyperonymes, coréférence, ou similarité grammaticale ou lexicale dans les thesaurus. Des chaînes lexicales peuvent être établies entre les phrases du document, entre le titre et le document ou la requête et le document. Les chaînes sont alors pondérées et les phrases sont jugées significatives lorsqu'elles sont traversées par un grand nombre de chaînes [Silber et McCoy, 2002].

Filtrage sémantique et linguistique

La méthode d'*Exploration Contextuelle* [Desclés, 1988] [Desclés, 1997] vise à identifier les connaissances linguistiques dans le texte en les restituant dans leurs contextes et en les organisant en tâches spécialisées. Le but de cette approche est la classification sémantique des phrases du document pour en extraire les informations pertinentes. Cette méthode utilise des marqueurs de divers phénomènes linguistiques qu'il est possible de repérer dans les textes [Desclés *et al.*, 1991]. L'approche développée est fondée sur la construction manuelle d'une base de données de marqueurs linguistiques et de *règles d'exploration contextuelle*. Ces règles appliquées aux phrases du texte source

vont filtrer les informations sémantiques indépendantes du domaine avec les étiquettes sémantiques hiérarchisées comme : énoncés structurants et thématiques, définition, causalité, soulignement, récapitulatifs, conclusion, etc. La stratégie de sélection des unités saillantes est fonction des besoins des utilisateurs (filtrage d'informations). Les segments de texte décorés avec les étiquettes choisies par l'utilisateur sont fournis en sortie.

Dans cette approche, d'abord il faut identifier une occurrence de marqueur sous la forme d'un *indicateur* répertorié et dans un deuxième temps, explorer le contexte de cette occurrence pour rechercher d'autres *indices* linguistiques, sous la forme d'occurrences d'indices complémentaires dans un espace de recherche. Un espace de recherche est un segment textuel (proposition, phrase, paragraphe...) déterminé à partir de l'indicateur, l'espace dans lesquels les indices complémentaires doivent être recherchés. L'exploration contextuelle est gouvernée par un ensemble de règles qui, pour un marqueur donné, recherchent d'autres indices explicites dans un espace de recherche déterminé par la règle, par conséquent la règle déclenchée peut attribuer une étiquette sémantique à un segment linguistique.

Cette méthode définit un langage de description qui permet au linguiste de constituer sa base de données linguistiques. Le linguiste déclare soit des formes lexicales significatives des indices ou des indicateurs qu'il organise en classes : par exemple le mot *présentons* déclaré comme un indicateur ce qui est placé dans la classe *Éverbe-présentatif* ou *qui suivent* déclaré comme un indice regroupé dans la classe *Épartie-document1*, soit des combinaisons de classes. Ces combinaisons permettent de déclarer des lexies ; par exemple, la déclaration de *il + &etre + &importance* permet au linguiste de déclarer des lexies du type *il est primordial ; il est particulièrement important ou il est, ..., essentiel*. Chaque règle comprend une partie **Déclaration d'un Espace de Recherche E**, une partie **Condition** et une partie **Action** qui n'est exécutée que si la partie **Condition** est vérifiée. La partie **Déclaration d'un Espace de Recherche E** permet de construire un segment textuel, l'espace de recherche, en appliquant différentes opérations sur la structure du texte construite par le moteur d'exploration contextuelle. La partie **Condition** explicite les conditions que doivent vérifier les indicateurs et les indices complémentaires. La partie **Action** attribue une étiquette à un segment textuel. La figure 2.1 présente un exemple de règle écrite dans ce langage. La règle de nom *t* est *Rthématique* attribuée à la tâche *Thématique* et déclenchée si un indicateur appartenant à la classe *Éverbe-presentatif*, qui regroupe des formes lexicales des verbes presentatifs, est présent dans l'espace de recherche *E1* construit avec les formes lexicales de la phrase dans lequel apparaît l'indicateur. Une première condition exprime qu'une occurrence d'un indice de la classe *Épartie-document1* et

Épartie-document3, qui regroupe certains mots comme lignes, qui suivent, etc., doit être présente dans la phrase et une seconde condition précise une contrainte sur l'agencement entre l'indicateur et l'indice, comme par exemple la présence possible d'une incise. La partie **Action** indique que l'étiquette *Annonce-Thématique* est attribuée à la phrase en question.

Nom de la règle : *Rthématique* ;
Tâche déclenchante : *Thématique* ;
Commentaire : capte un schéma du type : Dans les lignes qui suivent ... nous présentons ...
Classe de l'Indicateur : *Éverbe-presentatif* ;
E1 := Créer-espace(voisinage Indicateur)
L1 := *Épartie-document3*
L2 := *Épartie-document1*
Condition : Il-existe-un-indice x appartenant-à E1 tel-que classe-de x appartient-a (L1) ;
Condition : Il-existe-un-indice y appartenant-à E1 tel-que classe-de y appartient-a (L2) ;
Precede (x,y, contrainte) ;
Action : Attribuer(PhraseParent *Annonce-Thématique*)

FIG. 2.1 – Un exemple de règle écrite dans le langage formel

Le système ContextO (élément du projet d'ingénierie linguistique FILTEXT [Minel *et al.*, 2001]), conçu selon ces principes, utilise diverses classes de règles d'exploration pour identifier des étiquettes d'ouverture et de fermeture de thème ainsi que des relations causales.

L'exploitation de critères plus linguistiques est aussi représentée par le système de Lehman [Lehman, 1995] [Lehman, 1997], qui s'appuie sur le repérage de marqueurs de surface dans les phrases d'un document pour déterminer si une phrase est à retenir pour la construction du résumé. C'est la méthode utilisée dans le système commercial *Pertinence Mining* que nous utilisons au chapitre 7 de l'évaluation.

Théorie de la Structure Rhétorique (Rhetorical Structure Theory : RST)

Cette approche exploite les théories de discours pour trouver les relations entre les propositions des phrases (noyau et satellite) [Mann et Thompson, 1987]. Le but de la RST est de décrire les textes, plutôt que les processus qui sous-tendent leur création et leur interprétation. Elle postule un ensemble de possibilités de structures, divers types de *blocs de construction* dont on peut observer les occurrences dans les

textes. Ces *blocs* se situent à deux niveaux, le principal ayant trait à la nucléarité et aux relations. Le schéma le plus fréquent : deux segments de texte sont reliés de telle sorte que l'un d'eux joue un rôle spécifique par rapport à l'autre. Par exemple une affirmation suivie de la démonstration étayant cette affirmation. La RST postule une relation de *démonstration* entre les deux segments. Elle considère également que l'affirmation est plus importante pour le texte que la démonstration particulière, et marque cette préséance en dénommant le segment d'affirmation un noyau et le segment de démonstration un satellite.

RST se compose d'un nombre des relations rhétoriques qui relient des segments de textes, et d'un certain nombre de schémas récursifs indiquant comment les textes sont structurellement composés dans la représentation comme un arbre. La plupart des relations sont binaires et asymétriques : elles relient un noyau (central au but de l'auteur) et un satellite (moins central).

Les deux, [Ono *et al.*, 1994] et [Marcu, 1997] se sont servi de RST comme base pour le résumé des textes en construisant un arbre rhétorique basé sur la présence des marqueurs explicites de discours et en utilisant des règles heuristiques de décider du meilleur arbre rhétorique pour un texte donné. Dans le cas de Marcu, l'unité minimale de l'analyse est la clause tandis que pour Ono *et al.* c'est la phrase. Une fois l'arbre obtenu, des unités des textes doivent être extraites pour le résumé. Dans l'approche de Ono *et al.* des phrases sont pénalisées selon leur rôle rhétorique dans l'arbre. Un poids de 1 est indiqué aux unités satellites et un poids de 0 est donné aux unités de noyaux. Le score final d'une phrase est donné par la somme des poids de la racine de l'arbre à la phrase. Des phrases peuvent être ordonnées en ordre croissant des poids et être utilisées comme base pour la construction de résumé. Dans l'approche de Marcu, chaque noeud de parent identifie ses enfants nucléaires comme saillants, en permutant leurs enfants à leur niveau. Le processus est récursif vers le bas de l'arbre. Le poids d'une clause est donné par son niveau obtenu après la permutation. Les poids sont utilisés pour produire une liste ordonnée de clauses qui peuvent être utilisées comme base pour le résumé.

Résumé par apprentissage

La méthode basée sur l'apprentissage sur un corpus, par [Kupiec *et al.*, 1995] a mis en application un classificateur Bayésien qui calcule la probabilité d'une phrase dans un document source d'être incluse dans le résumé. Dans cette approche, afin d'entraîner le classificateur, on a utilisé un corpus de 188 paires de documents/résumés (écrits par

les abstracteurs professionnels) dans des domaines scientifiques. Les caractéristiques considérées des phrases étaient : longueur, expressions prototypiques (*cue-phrases*), position de la phrase dans le paragraphe, fréquences des mots de la phrase et mot en majuscule. En utilisant cette approche probabiliste, ils ont constaté que la position caractéristique qui donne la meilleure exécution est une combinaison de position, expressions, et longueur de la phrase.

Approches mixtes : Combiner les méthodes linguistiques et statistiques

Diverses méthodes ont utilisé une combinaison de différentes fonctions. Dans le système de [Teufel, 1998] cinq traits sont retenus : la présence des motifs lexicaux particuliers, la position de la phrase dans le texte, la longueur de la phrase, la présence de mots thématiques, la présence de mots dans un titre. Il construit des classes rhétoriques de phrases en augmentant les scores en fonction des traits présents dans les phrases. Les phrases ayant les meilleurs scores forment le résumé.

Une extension de ce modèle par [Mani et Bloedorn, 1998] est effectuée, avec des classificateurs Bayésiens, en utilisant des règles d'arbre de décision sur un corpus des articles comprenant le résumé de l'auteur. Le but est de construire deux formes de résumés : génériques et orientés par utilisateur. Cette expérience a obtenu de bons résultats sans utilisation des expressions prototypiques.

SUMMARIST [Hovy et Lin, 1999] utilise dix-huit caractéristiques différentes et leur combinaison à l'aide de l'approche d'apprentissage par machine C4.5 [Quinlan, 1989]. Hovy et Lin se concentrent sur l'identification des topiques et la fusion pour la production du résumé. Ils décrivent un processus en trois parties : identification du topique (correspondant à la phase d'analyse), interprétation (correspondant à la phase de transformation) et génération (correspondant à la phase de synthèse). Dans l'identification du topique, ils utilisent un nouvel algorithme pour l'identification automatique des positions de phrase comportant des topiques importants. De plus, ils combinent manuellement la position de phrase avec d'autres caractéristiques appropriées pour pondérer les phrases. Sur la fusion de topique, ils explorent le concept en utilisant la base de données de WordNet, la catégorisation des textes, et le regroupement de textes (*clustering*).

2.2.2 Génération automatique de résumé

Jing et McKeown [Jing et McKeown, 1999] montrent que des résumés humains sont souvent construits à partir du document source par un processus de *couper et coller* (*cutting and pasting*) de fragments de document qui sont alors combinés et régénérés comme des phrases de résumé. Un système peut être développé par génération de résumé, par le processus suivant : réduction du texte en supprimant les fragments sans importance, fusion et génération de l'information pour combiner les fragments extraits, c'est-à-dire une transformation syntaxique et réordonnancement des segments du texte. [Jing, 2002] décrit la construction d'un corpus représentant le procédé de *cut-and-paste* utilisé par des humains ; un tel corpus peut être utilisé pour l'entraînement d'un système de résumé automatique.

Certaines approches de génération de résumé utilisent les patrons/schéma. Le principe de ces approches est de prédéfinir des gabarits dont les slots spécifient les informations importantes, de remplir le gabarit à partir d'informations extraites à partir du document source et de générer le contenu du gabarit comme texte du résumé.

Le système SumUM, développé par [Saggion et Lapalme, 2002] produit de courts résumés automatiques de longs documents scientifiques et techniques par une technique de génération. Pour établir ce système, des résumés écrits par des professionnels ont été manuellement alignés avec les documents originaux pour identifier le type d'information à extraire. Étant donné la forme structurée des articles scientifiques, la majeure partie de l'information dans les résumés a été trouvée dans la première section de l'article (introduction) (40%) dans les titres ou les légendes (23%). À partir de ces observations, les auteurs ont développé une approche, appelée Analyse Sélective, qui produit des résumés indicatifs et informatifs. SumUM produit le résumé en deux étapes : l'utilisateur reçoit d'abord un résumé indicatif où ce dernier peut identifier les topiques du document qui l'intéressent ; le système génère ensuite un résumé informatif qui élabore les topiques choisis par l'utilisateur. L'implémentation est basée sur une analyse syntaxique et sémantique, l'identification conceptuelle et la régénération des textes. L'exécution de cette méthode se fonde sur la sélection des types particuliers d'information du texte source ; l'instanciation de différents types de gabarits ; la sélection des gabarits afin de produire un résumé indicatif ; la régénération d'un nouveau texte court qui indique les topiques du document et l'expansion du texte indicatif avec l'élaboration des topiques.

Sélection indicative : Le système considère les phrases qui ont été identifiées comme le transport de l'information indicative. Étant donné une phrase S et un type d'information T le système vérifie si la phrase correspond à certains des modèles as-

sociés au type T. Des modèles indicatifs contiennent des variables, des constructions syntaxiques, des concepts de domaine et des relations. Les concepts sont regroupés en catégories comme : les auteurs (les auteurs de l'article, leur affiliation, chercheurs), le travail des auteurs (le travail, l'étude), l'activité de recherche (la situation réelle, le besoin de la recherche, le problème, la solution, la méthode), l'article (le papier, le composant de papier), les objectifs (l'objectif, le centre), les activités cognitives (la présentation, l'introduction, l'argument). Les relations se réfèrent aux activités générales de l'auteur pendant la recherche et l'écriture du travail : l'étude (examine, étudie), le rapport du travail (le prix, le rapport), la motivation (l'objectif, le centre), la pensée (l'intérêt, l'avis), l'identification (définir, décrire).

Sélection informative : Pour chaque topique potentiel et phrase où il apparaît, le système vérifie si la phrase contient un marqueur informatif et satisfait un critère informatif. Si c'est le cas, on considère que c'est un sujet potentiel du document, un gabarit informatif sera instantié et une liaison sera créée entre le sujet et le gabarit qui fera partie du résumé informatif.

Génération : Le contenu indicatif tient compte de l'information positionnelle et de l'ordre conceptuel suivant : la situation, le problème, la solution, l'introduction d'entité, l'information d'actualité, le but de l'entité conceptuelle, le centre d'entité conceptuelle, des aspects méthodologiques, des inférences et l'information structurelle. Les gabarits du même type sont regroupés s'ils sont apparus dans l'ordre dans la liste. Les types considérés dans ce processus sont : le sujet du document, le sujet des sections et l'information de signalisation. Les gabarits triés constituent le plan du texte. Chaque élément dans le plan du texte est employé pour produire une phrase. La structure de la phrase dépend du type de gabarit. L'information sur la situation, le problème, le besoin de la recherche, est reprise de l'original avec peu de modifications.

2.2.3 Résumé automatique de multi-documents

Certaines approches ont été développées pour la production de résumés de documents multiples. Le but est d'identifier ce qui est commun et ce qui diffère dans une variété de documents reliés et d'enlever les informations répétitives du résumé [Mani, 2001].

Les trois problèmes principaux pour le résumé multi-documents décrit sont [Radev *et al.*, 2002a] :

- la reconnaissance des unités saillantes redondantes ;

- l'identification des différences entre les documents ;
- la cohérence du résumé même quand le contenu vient de différents documents sources.

La concaténation des résumés mono document pour chaque texte dans une collection n'est pas satisfaisante, à cause des informations répétitives. Les résultats d'évaluation de DUC 2003 [Over et Yen, 2004] montrent que le meilleur système de production de résumé mono document avec une méthode de concaténation simple peut donner un résultat faible en multi-documents.

En 2002, lors d'un workshop d'ingénierie linguistique de Johns Hopkins University une équipe de chercheurs a développé le système MEAD [Radev *et al.*, 2002b] pour la production d'un résumé multi-documents. Pour chaque groupe de documents reliés, un centroïde est produit spécifiant des mots-clés et leurs fréquences dans l'ensemble des articles sources. Pour un taux de compression donné, l'algorithme choisit les phrases avec le plus de mots-clés du centroïde. Ces phrases sont considérées comme le centre des topiques du groupe. Les travaux les plus récents sur le système MEAD utilisent une combinaison linéaire de trois caractéristiques pour classer les phrases des documents sources : le score du centroïde, le score de position (plus élevé pour les phrases débutant d'un document) et la longueur pondérant les phrases les plus longues.

En utilisant ces critères, les phrases sont triées par scores et ajoutées au résumé jusqu'à la longueur désirée. Pour éviter la répétition et s'assurer que les phrases sélectionnées ne sont pas similaires à une autre, une métrique de cosinus est calculée pour toutes les paires de phrases sélectionnées.

2.3 Discussion

Nous avons présenté différentes approches de construction de résumé automatique. Le but des approches de création de résumés par **extraction** est de dégager des phrases pertinentes en respectant l'ordre d'apparition des phrases dans le document, leur juxtaposition constitue le résumé. Les avantages du résumé par extraction est la simplicité de la méthode et l'utilisation de phrases déjà faites. Ceci évite le problème de la compréhension et le problème de la génération de texte. Les problèmes du résumé par extraction sont les extraits qui sont souvent incohérents dans le texte surtout quand le contexte est omis. La majorité des systèmes basés sur cette technique ne tiennent pas compte de la structure thématique du document. Ceci peut entraîner des ruptures entre les différentes unités extraites et produire des résumés moins cohérents.

Les méthodes de **génération** de résumés utilisent la technique de la reformulation ou la réduction du texte. Les avantages de ces approches sont de produire un résumé plus cohérent avec une meilleure possibilité de compression. Mais pour obtenir des informations précises sur les éléments du document, par exemple pour des articles scientifiques, la génération de texte portant sur les points importants (sujet, but, hypothèse, méthodologie, résultats et conclusion) exige des gabarits et des patrons complexes basés sur les connaissances du domaine.

Dans le prochain chapitre, nous décrivons les travaux en résumé de texte juridique ainsi que les productions manuelle et automatique de résumé dans ce domaine. Nous allons voir que ces travaux sont limités parce que les résumés manuels sont très coûteux et les approches automatiques sont restreintes. Dans le chapitre 4, nous présentons notre étude sur la comparaison des résumés manuels avec les textes sources afin d'étudier le processus d'analyse manuel des jugements. Cette étude nous conduit au développement une approche de production du résumé de textes juridiques que nous décrivons au chapitre 6. Dans notre méthode, nous considérons d'abord l'organisation du document afin d'en identifier la structure thématique et nous sélectionnons les phrases les plus saillantes dépendamment de l'importance d'un thème dans le texte. Le résumé produit par notre approche a l'avantage d'être plus cohérent et compréhensible par rapport aux résumés produits par des méthodes d'extractions actuelles parce que chaque phrase de résumé est associée avec son rôle thématique représentant le contexte dans le jugement source. Dans cette thèse, nous n'avons pas effectué de reformulation des phrases du résumé car les juristes exigent de conserver les phrases originales afin de minimiser la possibilité de mauvaise interprétation du texte du jugement.

Chapitre 3

Résumé de documents juridiques

Des milliers de professionnels utilisent l'information juridique et leurs ordres professionnels sont de plus en plus intéressés à répondre à ces besoins. D'autres travaillent à préparer ces produits d'information dans diverses maisons d'édition. D'autres encore, au gouvernement ou dans les universités, consacrent leurs énergies à augmenter l'offre gratuite des documents juridiques officiels.

Au Canada, le Centre de Recherche en Droit Public (CRDP) a mis sur pied le projet CanLII (Canadian Legal Information Institute) qui a pour objectif de créer une bibliothèque de droit virtuelle donnant un accès internet gratuit aux décisions de tous les tribunaux judiciaires canadiens. Le nombre de jugements rendus annuellement est de plus de 200 000. Cette abondance de textes juridiques sous forme numérique nécessite la création et la production d'outils informatiques performants en vue d'extraire l'information pertinente sous une forme condensée, avec des méthodes fiables et peu coûteuses.

Mais pourquoi s'intéresse-t-on au traitement des décisions juridiques passées, ainsi qu'à leurs résumés ? D'abord parce que pour régler un problème juridique dont la solution ne se trouve pas directement dans la loi, les avocats cherchent de la jurisprudence afin de trouver les cas semblables et d'analyser les solutions adoptées par les avocats des parties et les juges dans le passé. Car chaque décision tient lieu de *loi* entre les parties et en justifie la solution. Donc une jurisprudence constitue un *précédent* juridique puisqu'il est possible d'en extraire une règle de droit pouvant servir à disposer d'affaires semblables. Pour traiter un cas juridique il faut examiner des centaines de précédents qui sont très longs à étudier. Lire tous ces documents, pour trouver les décisions pertinentes pour cette affaire pouvant être fastidieux, les experts et les étudiants de droit sont demandeurs de résumés de décisions judiciaires.

Voici un exemple d'utilisation de précédent pour déterminer un cas qui n'est directement dans les lois. Les droits dans les deux premières situations suivantes sont clairement expliqués dans la loi mais la troisième situation est un cas plus difficile à déterminer :

- un employé quitte son emploi sans raison logique, il n'aura pas droit à l'assurance emploi.
- une compagnie réduit le nombre de ses employés à cause de problèmes économiques, il y aura dans ce cas droit d'assurance emploi pour les employés.
- Si un employé quitte son emploi à cause de *problèmes psychologiques* dans l'entreprise, aura-t-il droit à l'assurance emploi ?

Les *problèmes psychologiques* ne sont pas traités explicitement dans la loi, mais les avocats peuvent consulter la jurisprudence pour des cas semblables pour voir comment le juge a analysé et décidé. Par exemple, si le problème est causé par un harcèlement ou si le médecin recommande de changer d'emploi, les juristes peuvent déduire que ce sont des raisons valables pour quitter un emploi. Dans les cas flous, le juge peut ajouter son interprétation de la loi.

Le besoin d'accéder rapidement aux informations et d'extraire les idées principales des textes longs devient de plus en plus problématique. Sur le site de la bibliothèque de droit de l'Université de Montréal dans la rubrique *Ce mois-ci à la bibliothèque de droit- Questions et réponses sur l'index et résumés*, les étudiants demandent souvent la possibilité de savoir, si une décision donnée est repérable, sous forme de résumé dans les différentes bases de données. Les journalistes demandent souvent aux juges de publier des sommaires accompagnant leurs jugements pour faciliter les reportages par les médias.

Étant donné que plusieurs décisions sont disponibles en texte intégral dans les bases de données, on aimerait obtenir des résumés sous forme d'extraits condensés automatiques pour éviter la production manuelle de résumé qui peut être très coûteuse.

Au Québec, le Répertoire électronique de jurisprudence du Barreau (REJB) et la Société québécoise d'information juridique (SOQUIJ) sont deux organismes qui fournissent des résumés manuels pour les ressources juridiques, mais le temps et l'expertise nécessaires augmentent le coût d'utilisation de leurs services. Par exemple le prix d'un résumé de SOQUIJ et son texte intégral (en 2004) est 10\$ can. , alors que pour traiter un seul cas il faut souvent examiner des centaines de jugements comme précédents juridiques. Une requête sur un sujet, par exemple les langues officielles du Canada, dans la base de données de CanLII sur la collection des décisions de la Cour fédérale retourne 500 jugements comme précédents juridiques, d'où l'intérêt de la production

automatique de résumé.

La Société québécoise d'information juridique (SOQUIJ) décrit l'objectif de production de résumé par l'humain pour les ressources juridiques comme suit :

Le sommaire de l'arrêtiste poursuit deux objectifs : d'abord livrer l'essence du jugement clairement et avec concision pour permettre une consultation facile et rapide; et de fournir suffisamment d'informations sur le jugement pour permettre au lecteur de décider, en connaissance de cause, si celui-ci peut être pertinent à sa recherche.

Une raison de la difficulté de ce travail est la complexité du domaine : des vocabulaires spécifiques au domaine juridique et des interprétations d'expressions judiciaires peuvent amener des ambiguïtés; par exemple le mot *sentence* en anglais peut avoir deux sens différents : l'un est la phrase et l'autre est, dans un sens plus particulier du domaine, la condamnation ou la peine. Un autre exemple est le mot **dispositif** dont le sens courant est le mécanisme et la manière dont sont disposées les pièces d'un appareil alors que son sens en droit signifie la partie terminale d'un jugement dans laquelle est indiqué ce qui est décidé.

Les expériences des systèmes de résumé automatique actuels ont été essentiellement limitées à l'étude des articles journalistiques et à quelques systèmes sur les articles scientifiques [Saggion et Lapalme, 2002]. Il y a des différences importantes entre le langage journalistique et le langage juridique : statistique des mots, mots de titre et les relations de chaîne lexicale entre les mots du titre avec les idées clés du texte, relations entre une phrase avec sa précédente et la suivante, il en est de même pour les paragraphes et la structuration du texte. Pour les jugements, nous montrons dans cette thèse qu'il est possible de définir les structures discursives pour les différentes parties de la décision et de leur assigner des thèmes. Les articles journalistiques répètent souvent le message le plus important mais, en droit, le message est répété rarement plus d'une seule fois. Le traitement des textes juridiques demande une attention particulière et il n'est pas facile d'adapter les techniques déjà développées pour d'autres types de textes.

En considérant ce besoin de production des résumés de textes juridiques, nous allons élaborer une méthode de constitution automatique de fiches de résumés applicable au domaine juridique.

Dans ce chapitre, nous introduisons d'abord les notions de base de droit ainsi que les caractéristiques du domaine et du corpus. Par la suite nous décrivons l'état de l'art et les travaux pertinents sur la condensation automatique de texte et la problématique

du domaine.

3.1 Types de résumés juridiques

Voici quelques types de résumés de texte juridique, en fonction du contenu du texte, du but spécifié pour chaque résumé, et de la tâche demandée par l'utilisateur :

Résumé indicatif — Le résumé d'un texte juridique mentionnant brièvement tous les sujets contenus dans ce dernier.

Exemple — Les résumés créés par le Répertoire Électronique de Jurisprudence du Barreau (REJB) et la Société Québécoise d'Information Juridique (SO-QUIJ). Ils reprennent les textes intégraux de décisions publiés par les différentes cours judiciaires canadiennes comme la Cour d'appel, la Cour suprême du Canada, etc.

URL • Société québécoise d'information juridique : <http://www.soquij.qc.ca/>
 • Barreau du Québec : <http://www.barreau.qc.ca/>

Digest ou Fiche analytique — Ça extrait les éléments pertinents du texte de jugement et les dispose d'une façon à en faciliter la lecture.

Exemple — Les fiches analytiques construites par le Bureau du Commissaire à la magistrature fédérale. Ces fiches sont des résumés des décisions intéressantes, mais pas assez importantes pour être publiées au recueil en entier.

URL — Recueils de la Cour fédérale : <http://reports.fja.gc.ca/>

Sommaire (Headnote) — Le résumé des points juridiques pertinents déterminés par une cour. Les sommaires sont écrits par les arrêtiistes et permettent une consultation rapide du jugement. Le résumé de l'arrêtiiste est toujours placé en-dessous de l'intitulé de la décision et présenté en caractères italiques.

Résumés législatifs — Ils sont établis par la Direction de la recherche parlementaire de la Bibliothèque du Parlement. Ce sont les documents explicatifs sur les projets de loi, examinés par le Parlement. Il en existe un sur la plupart des projets de loi. Ils sont rédigés après l'étape de la première lecture, par des analystes qui connaissent bien le domaine des politiques ou du droit visé par le projet de loi.

URL — <http://www.parl.gc.ca/LEGISINFO/index.asp?Lang=F>

3.2 Ressources juridiques

Une ressource documentaire importante dans le domaine juridique est formée par les recueils de jurisprudence. La jurisprudence (*case-law*) réunit un ensemble des jugements, arrêts, décisions et avis publiés qui interprètent et précisent le sens des textes, les lois et règlements, par lesquelles les tribunaux supérieurs statuent sur des points de droit. Ces recueils, qui font l'objet d'annotations, ont été publiés dans des champs disciplinaires très divers : en matière civile, commerciale, administrative, etc. Leur présentation est le plus souvent chronologique.

Pour ce projet de recherche, deux grands corpus de jurisprudence étaient à notre disposition : le premier est la base des données *CanLII*¹, qui est une ressource permanente en droit canadien. Il diffuse des sources primaires du droit. Il est réalisé par l'équipe LexUM du Centre de recherche en droit public (CRDP) de l'Université de Montréal. Le deuxième est le *Recueil des arrêts de la Cour fédérale*², la Cour Fédérale applique du droit fédéral, qui est de type *common law*, c'est-à-dire que les *précédents* judiciaires- les décisions antérieures qui ont résolu un problème juridique- y jouent un rôle important comme source de droit. Le Bureau du Commissaire à la magistrature fédérale (BCMF) diffuse ces décisions de Cour fédérale ainsi que les résumés de certaines décisions.

Dans cette thèse, à la suggestion du CRDP, nous avons travaillé sur les corpus de décisions publiés de la Cour fédérale en raison du fait que ses décisions sont susceptibles d'intéresser des juristes provenant de toutes les provinces. Le fait d'avoir déjà des résumés modèles pour certaines décisions est aussi un avantage. Les décisions et les jugements délivrés par la cour fédérale, en général, sont bien structurés avec une rédaction judiciaire de bonne qualité. Ainsi, nous avons pu nous en servir comme un corpus modèle pour notre étude.

3.3 Travaux en résumé de texte juridique

Nous présentons deux types de création de résumé juridique : la production manuelle et les approches automatiques.

Dans notre recherche, le résumé manuel pourrait être utilisé comme référence, c'est pourquoi nous décrivons les résumés produits par deux organismes canadiens qui pro-

¹<http://www.canlii.org/>

²<http://reports.fja.gc.ca/>

duisent des résumés juridiques manuellement. Dans le chapitre 4, nous présenterons notre étude du corpus et l’alignement de résumés modèles et de textes sources pour en retirer les informations importantes.

3.3.1 Production de résumé manuel

Jurisprudence Express

Jurisprudence Express publie le résumé de tous les jugements motivés rendus par les cours de justice du Canada. Les résumés des décisions publiées, reprises en texte intégral dans le recueil, lequel contient également un plan de classification, une table de la législation citée, une table de la jurisprudence citée, une table de la doctrine citée, une table d’interprétation, un index, un suivi des appels ou révisions judiciaires des décisions publiées et une table de corrélation entre l’express et le recueil.

L’arrêteste relève dans son traitement les principales références faites dans le jugement à la loi et à la jurisprudence. Les coordonnées du jugement (nom des parties, nom du juge, date, district, etc.) suivent le sommaire.

AZIMUT est la banque en ligne qui contient les résumés et les textes intégraux des jugements rapportés au Jurisprudence Express. Comme la plupart des résumés juridiques, l’accès aux résumés est payant. La figure 3.1 montre l’exemple d’un résumé manuel créé par la Société Québécoise d’Information Juridique (SOQUIJ). Ce plan contient trois rubriques. Le premier champ expose les informations sur le jugement : les parties en litige, la juridiction (nom de cour, lieu, numéro de référence), nom du juge, date, indexation (la catégorie du sujet), et interprétation (les droits applicables sur le cas). Le deuxième champ, le résumé comporte un extrait sur l’analyse du juge et le troisième champ, la décision montre le résultat du jugement. Un résumé de SOQUIJ indexé avec son texte de la jurisprudence coûte 10\$ (tarification de 2004).

Fiches analytiques

Le Recueil des arrêts de la Cour fédérale rend disponibles toutes les décisions de la Cour fédérale destinées à être publiées dans le recueil officiel, dans leur intégralité ou sous forme de fiche analytique. Ces fiches analytiques sont des résumés des décisions intéressantes, mais pas assez importantes pour être publiées au recueil en entier. Il présente le contenu intégral du recueil officiel, anglais et français à compter de l’année 1993, publié dans les deux langues officielles. Ces décisions sont importantes pour des

Résumés SOQUIJ 2001/02/06

Résumé 379 mots, jugement 1514 mots

1. Informations sur le jugement

Parties — Gosselin c. Québec (Procureur général)

Jurisdiction — Cour d'appel (C.A.), Montréal, 500-09-001092-923

Décision de — Baudouin et Robert

Date — 1999-04-23

Références — AZ-99011340

Indexation — SOCIAL (DROIT)- sécurité du revenu - aide sociale - bénéficiaire de moins de 30 ans [...]

Interprétation — Charte canadienne des droits et libertés dans Loi de 1982 sur le Canada (L.R.C. 1985, app. II, no 44, annexe B, partie I), art. 7 [...]

2. Résumé

Appel d'un jugement de la Cour supérieure ayant rejeté une action en nullité, intentée par voie de recours collectif, de l'article 29 a) du Règlement sur l'aide sociale. Rejeté, avec dissidence.

L'appelante, qui représente tous les bénéficiaires de moins de 30 ans qui ont été assujettis à la disposition contestée entre le 17 avril 1987 et le 1er août 1989, soutient que celle-ci est inconstitutionnelle. Cette disposition avait pour effet, avant l'abrogation du règlement par l'adoption de la Loi sur la sécurité du revenu, entrée en vigueur le 1er août 1989, de réduire l'aide sociale versée aux bénéficiaires de moins de 30 ans, aptes au travail et vivant seuls. [...]

3. Décision

Mme la juge Mailhot : Le juge Iacobucci, dans l'arrêt *Law c. Ministre de l'Emploi et de l'Immigration du Canada*, énonce la démarche suivant laquelle il peut être décidé d'une atteinte au droit à l'égalité. Le programme d'aide accordé, envisagé dans sa globalité et dans son contexte, ne produit pas d'effets défavorables au sens de l'article 15 de la charte canadienne. Dans le cas contraire, l'article 1 de la charte rachèterait la mesure contestée. [...]

FIG. 3.1 – Exemple d'un résumé manuel écrit par SOQUIJ

Fiche analytique de la Cour fédérale

Résumé 209 mots, Jugement source 2 275 mots

Entre — Alli c. et Canada (Ministre de la Citoyenneté et de l'Immigration)

Dossier — IMM-1984-01

Référence neutre — 2002 CFPI 479

Juge — O'Keefe

Date — 26-4-02

Résumé — Demande de contrôle judiciaire d'une décision de la SSR selon laquelle le demandeur n'était pas un réfugié au sens de la Convention—Le demandeur, un citoyen nigérian, avait revendiqué le statut de réfugié au sens de la Convention en se fondant sur sa religion et sur son appartenance à un groupe social, à savoir les personnes contraintes à participer à des meurtres rituels et à des pratiques néfastes—Le demandeur avait refusé d'assumer le poste que son père occupait à titre de chef Oluwo Apeno, soit un culte dans le cadre duquel on se livrait à des sacrifices humains ; il avait été déshérité par sa famille et s'était converti au christianisme—On avait dit au demandeur que son peuple serait béni par les dieux s'il était tué et si son sang était utilisé pour d'autres sacrifices—Le demandeur s'est enfui—La SSR avait conclu qu'il n'existait aucune possibilité sérieuse que le demandeur soit persécuté s'il retournait au Nigéria et que l'État pouvait assurer à ce dernier une protection adéquate—Demande accueillie—La Commission a commis une erreur susceptible de révision au sujet des faits se rapportant à la protection étatique étant donné que la preuve documentaire montrait que la protection étatique ne pouvait pas être obtenue de la police.

FIG. 3.2 – Une fiche analytique d'une décision de la Cour fédérale

juristes de toutes les provinces du Canada, parce qu'ils représentent un précédent pouvant servir à disposer des cas semblables. La figure 3.2 montre un exemple du résumé qui est construit par extraction manuelle des segments pertinents du texte de décision de la Cour fédérale. Ce type de résumé est un bon exemple du résumé souhaité pour les juristes qui va nous servir comme référence.

3.3.2 Approches automatiques de production de résumé

Quelques approches sont proposées afin de fournir les outils permettant d'aider les avocats et les juristes, comme des compagnies privées QuickLaw³ au Canada et WESTLAW⁴ et LEXISNEXIS⁵ aux États Unis ; toutefois aucun ne satisfait complètement les exigences spécifiques de ce domaine d'application. Pour développer un système de traitement du texte juridique, il faut une équipe composée d'informaticiens et de juristes pour analyser le problème.

Dans cette section nous présentons quelques modèles proposés pour produire automatiquement des résumés de textes juridiques.

FLEXICON, University of British Columbia

FLEXICON [Smith et Deedman, 1987][Gelbart et Smith, 1991] [Smith *et al.*, 1995] est un système développé pour la gestion des informations juridiques et la production du résumé qui combine le traitement du texte avec raisonnement à base de cas. Cette approche utilise des modules d'extraction pour identifier les concepts, les cas, les législations, les faits et leurs relations dans la décision, afin de construire un profil structuré de document et produire automatiquement un sommaire (headnote). Les concepts sont identifiés par unification des mots du texte avec une liste d'expressions significatives, en appliquant des règles heuristiques simples.

FLEXICON génère des représentations de structures de documents juridiques. Le système effectue les opérations suivantes :

- extraction des informations d'en-tête sur le cas (case header) : les parties, la date, la cour, la juridiction, et des juges ;
- classification du sujet de la loi ;
- identification des paragraphes clés contenant l'essence de l'opinion de la cour.

Pour présenter les parties importantes du texte, le système génère une liste de quatre aspects juridiques : des concepts les plus significatifs, des faits, des cas et de l'ensemble des lois appliquées. Il calcule les poids de cette liste par ordre décroissant. Il extrait les paragraphes importants au complet et il élimine les paragraphes très courts et ceux qui contiennent les *citations* moins importantes dans le jugement.

Pour notre travail, malgré la simplicité de cette approche, le profil du document pro-

³<http://www.quicklaw.com/>

⁴<http://www.westlaw.com/>

⁵<http://www.lexisnexis.com/>

posé nous intéresse, car ce système a été développé pour les textes juridiques canadiens de type *common law* qui est notre corpus de texte.

SALOMON, Katholieke Universiteit Leuven

Le projet SALOMON [Uyttendaele *et al.*, 1996][Moens *et al.*, 1996] produit le résumé automatique de cas criminels belges (écrits en néerlandais). Ce système extrait les unités importantes des documents à partir du texte de jugement pour former un résumé. Le but est d'identifier et d'extraire les informations importantes à partir de la jurisprudence.

Deux méthodologies ont été utilisées pour développer SALOMON. D'abord il identifie la catégorie de cas, la structure et les unités non pertinentes des textes. Ce processus est basé sur la représentation des connaissances réunie dans une grammaire de texte. Ensuite, des données générales et fondamentales du sujet de la décision sont extraites. Deuxièmement, le système produit un résumé informatif des unités textuelles de l'opinion de la cour en utilisant les techniques statistiques.

Dans ce projet les connaissances linguistiques plus profondes sont utilisées. Il extrait les concepts et les unités textuelles saillantes grâce à l'identification des *cue words*, segments indicateurs et patrons de contexte développé en néerlandais. Cette recherche montre aussi l'intérêt d'utiliser la structure du discours des textes juridiques.

SUM, University of Edinburgh

C'est un projet en cours de développement [Grover *et al.*, 2003][Grover *et al.*, 2004] afin de produire les résumés des jugements de *House of Lords*, la cour suprême de l'Angleterre en utilisant les informations rhétoriques des phrases. Il est basé sur l'approche de [Teufel et Moens, 2002] proposée pour les articles scientifiques. SUM résume les textes juridiques en deux étapes :

- Décider quels rôles rhétoriques sont importants dans le texte source et lesquels sont utiles dans le résumé.
- Dans une collection de textes pertinents, décider pour chaque phrase, le rôle rhétorique qui la décrit mieux.

Un échantillon de 40 jugements de *House of Lords* ont été annotés comme ressources. Pour classifier les phrases, les classificateurs statistiques comme C4.5 et SVM sont utilisés. L'idée est de placer les phrases extraites par rapport aux sept rôles rhétoriques

(*fact, proceedings, background, framing, disposal, textual and other*), ce qui peut servir pour le résumé.

Dans cette approche, chaque phrase reçoit une étiquette comme son rôle rhétorique mais la classification des phrases reste à la limite de surface du texte, c'est-à-dire les phrases sont étiquetées selon leurs formes syntaxiques. La relation sémantique entre une phrase et le segment thématique qui la contient, dans le plan organisationnel du jugement, n'est pas considérée.

3.4 Discussion des méthodes de génération de résumé

Alors que nous avons le problème de grandes quantités de textes juridiques et le besoin de les présenter sous forme d'un résumé court, notre recherche montre qu'il n'y pas eu beaucoup de travail sur ce domaine et que le problème du traitement du texte légal n'est pas encore résolu.

Différents systèmes ont été développés pour différentes langues, comme l'allemand, le hollandais, l'anglais ou le français, mais une approche qui peut être efficace pour identifier les indicateurs marquant les phrases importantes dans une langue ne sera peut-être pas aussi utile pour d'autres langues avec d'autres types de styles. Il en est de même pour l'organisation du jugement qui peut différer selon les lois et la tradition juridique de chaque pays.

Les différentes méthodes utilisées pour traiter des textes juridiques montrent qu'une diversité de systèmes sont utilisés pour les tâches d'indexation, de catégorisation, et de production de résumés. La création automatique de résumé, semble la plus compliquée et demande une étude approfondie du sujet. Les différentes approches proposées essayent de régler une partie du problème mais le résultat d'évaluation des systèmes et la qualité des résumés produits n'est pas encore à un niveau satisfaisant. Au point de vue opérationnel, les résumés produits par les systèmes ont beaucoup de problèmes de cohérence et de lisibilité. Une des raisons de ce problème est le fait d'analyser les phrases indépendamment de leur contexte. Dans les jugements, un paramètre très important pour une interprétation correcte d'une phrase est de considérer le thème de discours dans lequel la phrase est située. Nous allons décrire dans le chapitre suivant notre analyse de corpus juridique afin d'identifier la structure organisationnelle des jugements et repérer les phrases saillantes en fonction des thèmes.

Chapitre 4

Étude de corpus juridique

4.1 Corpus d'étude

4.1.1 Composition

Notre corpus est composé de 3500 documents de jurisprudences en anglais, rendues par la Cour fédérale du Canada du tribunal de première instance des années 2000 à 2003, qui sont disponibles en format HTML sur le site CanLII¹. Nous avons analysé 50 jugements et leur résumés rédigés manuellement par un arrêviste, un résumeur professionnel. Ces jugements ont été sélectionnés par un avocat de CanLII comme représentatifs des différents sujets du domaine juridique avec des structures de textes qu'on peut considérer comme standard. Un nombre limité de résumés de décisions publiées de la Cour fédérale sont disponibles sous le nom de *fiches analytiques*².

Nous avons aussi étudié deux collections de documents, en français, regroupant 15 jugements et leurs résumés, produits par le Répertoire électronique de jurisprudence du Barreau (REJB), et le Recueil des arrêts de la Cour fédérale. Contrairement au projet SALOMON [Uyttendaele *et al.*, 1996] qui portait sur des cas criminels, notre recherche couvre différentes catégories de jurisprudences comme : accès à l'information, assurance-emploi, concurrence, douanes et accise, droit administratif, droit aérien, droit d'auteur, droit constitutionnel, droits de la personne, droit maritime, enquêtes, environnement, langues officielles, télécommunications, etc.

Afin d'estimer la taille moyenne des jugements devant être considéré par notre

¹<http://www.canlii.org>

²<http://reports.fja.gc.ca>

système, nous avons calculé la distribution des mots des 3500 documents de notre corpus. Nous observons que 75% des décisions ont entre 500 et 4000 mots (2 à 8 pages). Donc dans notre travail, nous concentrons nos travaux sur les textes dans cet intervalle de longueur ; 10% des documents ont moins de 500 mots, environ une page, et il n'est donc pas nécessaire de les résumer alors que seulement 15% des documents comportent plus de 4000 mots.

Dans notre corpus, la taille moyenne des jugements est de 3600 mots et la taille moyenne des résumés est 360 mots, c'est à dire un taux de réduction de 10%.

4.1.2 Structuration des textes juridiques

Dans certains domaines spécifiques comme le domaine juridique, le contexte peut changer la valeur sémantique d'une phrase. Souvent pour interpréter une phrase, il faut considérer sa place dans le fragment textuel et la situation dans laquelle cette phrase est employée. Pour la tâche d'extraction des phrases pertinentes de décisions de justice, il est aussi important de présenter les circonstances dans lesquelles les phrases ont été prononcées. Par exemple pour obtenir une interprétation correcte de la phrase *The application is dismissed*, il est nécessaire de prendre en compte son contexte : si cette phrase apparaît dans la partie du texte qui explique les histoires judiciaires sur le cas, elle signifiera une demande de révision sur une décision précédemment rendue par une autre cour de justice, mais si cette phrase apparaît dans la partie terminale du jugement, elle exprimera la décision finale du jugement en cours.

Pour aborder ce problème, nous avons étudié l'organisation générale de textes de jurisprudences. D'après nos analyses, les jugements sont organisés selon une macro-structure qui contient différents niveaux d'informations, indépendamment du domaine de jugement. Les travaux expérimentaux de la juge Mailhot de la Cour d'appel du Québec [Mailhot, 1996], utilisés pour guider les juges dans la rédaction d'un jugement, renforcent cette idée qu'il est possible de définir une structure organisationnelle pour ces décisions. Les jurisprudences sont organisées par le discours lui-même, ce qui permet de les segmenter en s'appuyant sur l'organisation discursive des contenus. L'ensemble des unités textuelles qui traitent du même sujet forme un segment thématique. Pour chaque thème, il existe une borne qui signale un changement dans le discours ou dans le thème du segment. Pour déterminer les bornes de segments thématiques, nous avons étudié plusieurs éléments indicateurs comme les titres significatifs des sections, les positions des fragments textuels et les expressions linguistiques que nous expliquerons en détail dans le chapitre 6. À l'intérieur d'un segment thématique, les phrases portent

sur un même sujet et chaque phrase est influencée par ses voisines.

Pour notre analyse de corpus, nous avons comparé les résumés modèles créés par un humain avec les textes des jugements originaux. Nous avons identifié la structure organisationnelle pour le jugement. Les paragraphes qui traitent d'un sujet sont considérés comme membres d'un groupe thématique. Nous avons annoté les segments avec une étiquette indiquant leurs thèmes. Nous avons aussi manuellement annoté les unités de citation, des unités textuelles (phrase ou paragraphe) données par le juge comme référence (par exemple à un article de loi). Les segments de citation occupent une taille considérable dans le jugement, mais ils ne sont pas considérés importants dans le résumé, donc ces segments seront éliminés lors des filtrages d'information.

Les unités textuelles considérées importantes par les résumeurs professionnels ont été alignées manuellement avec un ou plusieurs éléments du texte source. La table 4.1 montre un exemple de cet alignement entre un résumé humain et le texte intégral original. Nous y cherchions les relations entre les informations considérées importantes dans les résumés des arrêtistes et les informations dans les documents sources. Nous avons constaté que les arrêtistes produisent un résumé par extraction des unités importantes tout en suivant des lignes directrices du texte. L'assemblage de ces unités sélectionnées constitue le résumé.

L'identification des structures thématiques sépare les idées clés des détails secondaires d'un jugement et améliore la lisibilité du résumé en produisant des textes plus cohérents. Notre hypothèse est que, malgré la variabilité des catégories des jugements, on peut distinguer une structure sur les informations présentées dans une jurisprudence. Afin de repérer les thèmes du jugement, nous avons développé un segmenteur thématique basé sur des connaissances linguistiques et juridiques. Nous en dégageons un plan d'organisation thématique dans lequel les unités composant le discours prennent place au fur et à mesure de leur apparition dans le jugement. Nous allons présenter la fonctionnalité des thèmes et leur importance dans le jugement du point de vue des idées clés et principales.

Pour la structuration thématique, nous avons divisé un jugement en sept thèmes (table 4.2), ce qui est validée par les juristes : DONNÉES DE LA DÉCISION, INTRODUCTION, CONTEXTE, SOUMISSION, ISSUES, RAISONNEMENT JURIDIQUE et CONCLUSION. La présentation du résumé final respectera cette organisation afin de constituer une fiche de résumé de décisions en sept thèmes :

1. DONNÉES DE LA DÉCISION présentent la référence complète de la décision et la relation entre les parties sur le plan juridique : nom de la cour, lieu de l'audience,

Décisions de la Cour fédérale du Canada entre :
GUIDES LTD., et JASPER NATIONAL PARK PROFESSIONAL RIVER OUTFIT-
TERS ASSOCIATION, demandeurs et
LE PROCUREUR GÉNÉRAL DU CANADA, défendeur

Texte intégral Dossier : T-1557-98	Résumé de l'arrêtiste	Thème de section
[1] This application for judicial review arises out of a decision (the Decision) announced on or about the 30th of June 1998 by the Minister of Canadian Heritage (the Minister) to close the Maligne River (the River) in Jasper National Park to all boating activity, beginning in 1999.	Judicial review of Minister of Canadian Heritage's decision to close Maligne River in Jasper National Park to all boating activity beginning in 1999 to protect habitat of harlequin ducks-	Introduction
[7] The applicants offer commercial rafting trips to Park visitors in this area each year from mid-June to sometime in September.	Applicants offer commercial rafting trips on River.	Faits
[10] Consequently, a further environmental assessment regarding commercial rafting on the Maligne River was prepared in 1991. The assessment indicated that rafting activity had expanded since 1986, with an adverse impact on Harlequin ducks along the Maligne River.	1991 environmental assessment indicating rafting having adverse impact on harlequin ducks along river.	Faits

TAB. 4.1 – Alignement des unités textuelles du texte original du jugement avec les unités du résumé rédigé manuellement

- date du jugement, numéro de greffe, référence neutre, identification des parties, intitulé du jugement, autorités et doctrines citées.
2. INTRODUCTION contient les paragraphes explicatifs placés en tête du jugement qui présentent le sujet. Ils décrivent brièvement la situation qui se présente au tribunal et répondent aux questions **qui ? a fait quoi ? à qui ?**.
 3. CONTEXTE contient les faits et l'histoire judiciaire et l'ensemble des circonstances dans lesquelles s'insèrent des faits. Il recompose l'histoire du litige à partir des faits et des événements relatés lors de la présentation de la preuve et retenus dans le jugement.
 4. SOUMISSION présentent le point de vue d'une partie sur le problème.
 5. ISSUES (ou les questions de droit) identifient le problème juridique dont le tribunal

est saisi.

6. **RAISONNEMENT JURIDIQUE** discute à partir des contextes et des faits du litige en s'appuyant sur des références juridiques et des autorités afin d'en arriver à une conclusion. L'analyse du juge comprend deux parties principales. La première partie est la détermination des droits applicables pour le cas (potentiel). La deuxième partie porte sur l'application des droits sur les faits (actualisation). Il répond aux questions de droit et comporte une explication sur les motifs du juge. Cette partie est la plus importante du résumé d'une décision, puisqu'elle contient la justification de la décision finale de la cour et transmet la solution. Les motifs du tribunal doivent être aussi la réponse aux questions de droit soulevées par les parties. Après un jugement, les motifs deviennent règle de droit.
7. **CONCLUSION** est la dernière partie du procès qui fait connaître la décision du juge. Elle prononce le **dispositif**, la partie terminale du jugement dans laquelle est indiqué ce qui est décidé et les montants adjugés s'ils existent. Par exemple en droit pénal, il faut spécifier si la personne a été condamnée ou acquittée.

Dans notre corpus, nous avons identifié cette organisation textuelle sous forme de segments thématiques. En suivant les thèmes, le lecteur est guidé de la définition du problème en litige jusqu'à ce que la cour a retenu comme le résultat, qui peut soit arriver à une solution aux problèmes entre les deux parties, soit renvoyer le dossier à une autre cour pour compléter le processus juridique.

4.2 D'où viennent les informations

Dans le plan d'un jugement, nous avons considéré sept thèmes principaux (table 4.2). Il reste à déterminer parmi ces thèmes lesquels sont utiles pour le résumé avec quels degrés d'importance.

Selon nos observations, quatre thèmes jouent les rôles principaux : **INTRODUCTION**, **CONTEXTE**, **RAISONNEMENT JURIDIQUE** et **CONCLUSION**. La présence de ces quatre thèmes dans le jugement et dans le résumé est obligatoire. Dans la structure de résumé, nous préservons ces quatre thèmes et nous extrayons les phrases qui appartiennent à ces thèmes. Le thème **ISSUE** est optionnel dans le jugement, s'il existe nous l'ajoutons dans le résumé.

Afin de mesurer le degré de l'importance des quatre thèmes principaux du jugement, au cours de notre analyse de corpus de résumés alignés avec les textes de jurisprudences,

Structure thématique	Explications
DONNÉES DE LA DÉCISION	Nom de la cour de décision Lieu de l’audience Date du jugement Numéro de greffe Référence neutre Identification des parties Intitulé du jugement Autorités et doctrines citées
INTRODUCTION	Qui ? A fait quoi ? À qui ?
CONTEXTE	Faits recompose l’histoire du litige Histoire judiciaire
SOUSSION	Points de vue des parties sur le problème
ISSUES	Questions de droit à répondre pendant le jugement
RAISONNEMENT JURIDIQUE	Analyse du juge et détermination des faits Expression des motifs de la solution retenue
CONCLUSION	Décision finale de la cour

TAB. 4.2 – Structure thématique d’une jurisprudence

nous avons mesuré la distribution de l’information (en nombre de mots) pour chacun de ces thèmes. Les volumes occupés par ces champs dans le document source et le résumé sont données à la table 4.3. On y voit l’importance et la contribution de chaque thème de jurisprudence dans le résumé, ce qui nous servira à attribuer un score comme valeurs sémantiques des thèmes. Ces valeurs sémantiques augmenteront ou diminueront la chance des phrases candidates d’être sélectionnées dans le résumé final. Par exemple, à l’étape finale de contrôle de la taille de résumé, nous préférons plutôt une phrase du segment RAISONNEMENT JURIDIQUE qu’une du segment CONTEXTE, car le premier est jugé plus important à cause de sa valeur sémantique plus élevée.

Un jugement fait souvent référence à d’autres jurisprudences en les citant. Ces citations peuvent être placées en plusieurs segments thématiques selon les cas. Les phrases citées n’expriment pas les idées clés des jugements, mais les résultats retenus de ces citations sont importants. Habituellement, les citations contiennent deux phases, la première comporte les prétentions et arguments des parties, qui présentent le point de vue d’une partie sur le problème. Cette partie contient les soumissions des parties et leurs positions dans le litige que nous avons considéré comme le thème SOUSSION. La seconde phase, la citation des articles de lois, traite des prétentions en droits applicables sur le cas. Ces segments cités sont les sections ou les paragraphes des lois, les jurisprudences ou les doctrines que le juge utilise comme références dans son rai-

Structures thématiques	Jugement	Résumé
INTRODUCTION	5%	12%
CONTEXTE	24%	20%
RAISONNEMENT JURIDIQUE	67%	60%
CONCLUSION	4%	8%
Total	100%	100%

TAB. 4.3 – Les pourcentages de la contribution de chaque thème dans les résumés et les jugements originaux

sonnement. Dans plupart des cas, les citations sont considérées comme étant du thème RAISONNEMENT JURIDIQUE. Dans notre système, nous identifions les phrases citées, pour en identifier les résultats afin de les garder dans la liste des unités candidates à extraire.

4.3 Dictionnaire de connaissances linguistiques

Nous avons construit un dictionnaire de connaissances linguistiques et de vocabulaire spécifique au domaine juridique. Le dictionnaire contient 250 marqueurs linguistiques significatifs que nous avons observés dans le corpus. Ce dictionnaire est organisé en trois catégories. La première contient des marqueurs communs à plusieurs types de texte et elle est divisée en trois classes : les verbes, les concepts (noms, adjectifs et adverbes) et des indices complémentaires comme les marqueurs typographiques (ex. deux-points dans citation), prépositions (ex. *at page*), numéros (ex. *subsection 20*), et subordonnés relatives (ex. *that*). La deuxième catégorie de marqueurs contient les expressions prototypiques (*cue-phrases*) qui expriment les parties importantes du discours (ex. *finally*). La troisième catégorie contient les vocabulaires et les expressions spécifiques au domaine juridique qui permettent de désambiguïser les termes avec différents sens dans un dictionnaire général mais qui ont, en droit un sens particulier, par exemple la différence entre un *appel* téléphonique et un jugement porté en *appel*.

Nous avons écrit 50 classes des règles sémantiques afin de contrôler les conditions sur les présences et les positions des marqueurs de notre dictionnaire. Nous expliquons ces règles au chapitre 6.

4.4 Conclusion

Dans ce chapitre nous avons présenté le résultat de notre analyse de corpus afin d'identifier la structure thématique des jugements. Ceci est la base de notre méthodologie de la construction de résumé automatique.

Notre analyse montre la difficulté d'appliquer des systèmes de production des résumés génériques sur les domaines spécifiques comme traitement des documents juridiques, c'est-à-dire construire les résumés sans tenir compte des besoins des utilisateurs et des domaines d'application. Dans notre méthode, nous avons considéré la structure thématique du jugement comme le facteur le plus important. L'importance d'une phrase est relative à son thème.

Dans le prochain chapitre, nous décrivons notre méthodologie et la concrétisation de notre approche afin de développer un système opérationnel, comme un outil pour la représentation et la manipulation de la structure du document. Dans la présentation du résumé final, nous conserverons les rubriques essentielles de la constitution de fiche de résumé.

Chapitre 5

Méthode de constitution automatique de fiches de résumé

5.1 Introduction

Notre étude de corpus, basée sur une analyse des phrases des résumés modèles alignées manuellement avec les jugements sources, fait ressortir la structure thématique d'un jugement. Notre approche de résumé automatique est fondée sur cette analyse. Elle regroupe les paragraphes qui traitent du même sujet, ce qui conduit à détecter les frontières entre les différents thèmes du document. Le découpage du document divise le texte en segments thématiques. Pour chacun, nous attribuons un thème comme étant le sujet traité par les paragraphes du segment. D'après le thème du segment, nous extrayons des phrases saillantes qui contiennent les informations pertinentes sur ce thème. Mais avant l'étape de sélection, un filtrage des segments cités (jugés non importants) sépare les phrases essentielles du jugement aux phrases citées comme références. L'ensemble des phrases extraites constitue le résumé.

Dans les sections suivantes, nous expliquons notre méthode de production de résumé automatique qui repose sur l'identification de la structure thématique du jugement, en utilisant la technique d'extraction des unités saillantes, avec une présentation du résumé final sous forme d'une fiche contenant des rubriques homogènes d'informations. Cette fiche présente les informations considérées importantes associées à des thèmes précis, ce qui en facilite la lecture et la navigation entre le résumé et le jugement source. Pour chaque phrase du résumé produit, l'utilisateur peut en déterminer le sujet en regardant le thème associé à son segment thématique. Si une phrase semble plus importante pour l'utilisateur et qu'il désire plus d'information sur ce sujet, on peut lui

proposer le segment thématique entier contenant la phrase sélectionnée, pour obtenir les informations complémentaires sur le sujet. Dans notre système, il y a d'abord un prétraitement du texte, ce qui implique la séparation du document en unités textuelles (paragraphe, phrases, mots, nombres et ponctuations) et modules linguistiques (voir chapitre implémentation section 6.4). La constitution de la fiche de résumé se fait en quatre étapes (figure 5.1) : segmentation thématique, filtrage des unités moins importantes comme les citations des articles des lois, sélection des unités textuelles candidates pour le résumé et production du résumé selon la taille demandée. L'étape de segmentation repère le squelette structurel du texte. Chaque niveau de cette structure a un thème particulier dans le jugement. Selon le thème, nous cherchons à identifier les unités pertinentes dans les segments correspondants, tout en respectant la limite de la taille du résumé.

5.2 Étapes de la constitution de la fiche de résumé

5.2.1 Analyse thématique

Dans la première étape, nous nous intéressons à déterminer l'organisation du document original et la segmentation thématique ayant pour objet le découpage des textes en segments thématiquement homogènes.

Expérimentation

Pour la segmentation thématique, nous avons fait quelques expérimentations avec deux segmenteurs décrits par Hearst [Hearst, 1997] le système *TexTiling* et le segmenteur *C99* décrit par Choi [Choi, 2000] afin d'aligner les bornes de segment sur des frontières des découpages thématiques du texte. L'algorithme de *TexTiling* fonctionne d'après une fenêtre glissante qui se déplace sur le texte à segmenter. Pour chaque position du texte ainsi parcourue, il compare les deux parties de la fenêtre situées de part et d'autre de cette position à l'aide d'une mesure de similarité s'appuyant sur les mots présents dans la fenêtre. Les valeurs de cette mesure pour l'ensemble du texte permettent de situer les zones de faible cohésion lexicale. Celles-ci montrent les ruptures thématiques et donnent lieu à la définition de segment thématique. Hearst a constaté que pour des corpus de journal une fenêtre de 6 phrases, composée de 20 mots chacune, semble optimale. La similitude parmi deux fenêtres est calculée avec une similarité de cosinus. La valeur réelle de la similitude n'est pas utilisée pour calculer des ruptures

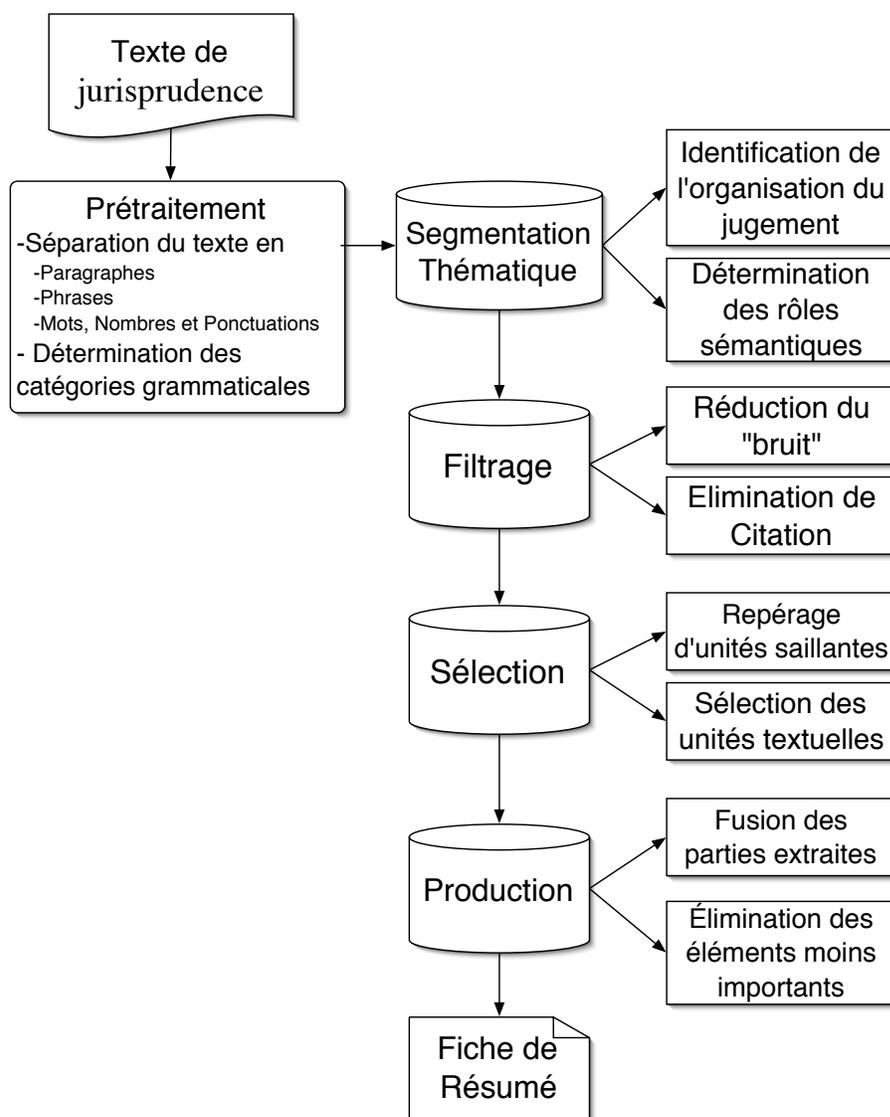


FIG. 5.1 – Les étapes de la constitution de la fiche de résumé

thématiques, l'algorithme ne considère que des différences relatives.

Dans ce travail, nous considérons également *C99*, un algorithme pour la segmentation linéaire des thèmes. Elle diffère de *TextTiling* parce qu'elle prend de vraies phrases comme unité de base et emploie une combinaison de la similitude (calculée parmi des phrases) pour le grouper. Après des étapes de prétraitement (élimination des mots vides et lemmatisation), l'algorithme calcule une matrice de similitude phrase par phrase, basé sur la similitude habituelle du cosinus. Puis, un arrangement de poids est appliqué à la matrice de similitude, afin de faire ressortir les différences dans la similitude des phrases. Une fonction de *clustering* (classification) hiérarchique est finalement appliquée pour diviser le document par thème.

Les résultats de ces segmenteurs numériques ne furent toutefois pas satisfaisants pour déterminer les structures thématiques des jugements. *TexTiling* nous donnait un segment thématique par paragraphe et *C99* ne nous donnait que quelques découpages entre les paragraphes et les citations. Dans un jugement, pour présenter une interprétation juridique sur les droits applicables et en déduire les solutions possibles pour le litige, le juge cite les références et il présente les paragraphes de lois tels qui sont documentés en anglais et en français. Ceci produit un texte avec des découpages thématiques irréguliers par rapport un texte journalistique, avec des changements de langues dans le texte. Par exemple, dans l'analyse pour justifier sa décision, le juge cite souvent les références ; il y a donc deux changements thématiques dans le segment : la transition de l'analyse vers les citations et la transition de citation vers la suite de l'analyse, alors que la sémantique du segment n'est pas changée et que le sujet est toujours le raisonnement du juge.

Nous avons donc procédé à une segmentation thématique basée sur des connaissances spécifiques au domaine juridique. Pour y arriver, il faut identifier les bornes de la structure thématique pour encadrer les segments du texte associés avec un thème précis dans la jurisprudence.

Dans l'exploitation de la structuration du jugement, nous analysons les citations dans les textes. La détection des citations est importante pour la prochaine étape de filtrage d'information. Afin d'identifier les citations dans les jugements, nous avons développé un module qui identifie les marqueurs directs et indirects de la citation. Nous décrirons ce processus dans la section 5.2.2.

Module de segmentation thématique

Afin d'extraire les idées essentielles, nous avons cherché les thèmes dans la structure d'un jugement qui désignent les valeurs sémantiques des phrases. Un thème peut changer l'importance d'une phrase selon différentes interprétations. Par exemple, l'interprétation sémantique de la phrase suivante *the plaintiff should be awarded* dépend de sa position dans le texte. Pour une même forme syntaxique de phrase, il y a deux interprétation possibles : si cette phrase est située dans le thème CONCLUSION alors elle sera une phrase importante qui annonce la décision finale du jugement, mais si elle est située dans un autre thème signifiera simplement une condition : *the plaintiff should be awarded only where there is a serious issue for consideration by the Court*.

Pour désambiguïser des phrases et obtenir les interprétations correctes des phrases, il est nécessaire d'identifier la structure du texte et les segments thématiques.

Catégories des titres	Marqueurs linguistiques	Exemples des titres
INTRODUCTION	<i>introduction, summary</i>	<i>Introduction, Summary</i>
CONTEXTE	<i>facts, background</i>	<i>The factual background, Agreed statement of facts</i>
SOUMISSIONS	<i>position, argument, submission</i>	<i>Respondent's position, Arguments of party, Applicant's submissions</i>
ISSUES	<i>issue, question of law</i>	<i>Issues, The questions of law</i>
RAISONNEMENT JURIDIQUE	<i>analysis, decision, discussion</i>	<i>Analysis and Decision of the court</i>
CONCLUSION	<i>conclusion, disposition, cost</i>	<i>Conclusion and Costs</i>

TAB. 5.1 – Les marqueurs significatifs qui se trouvent dans les titres des sections

Notre segmenteur thématique est fondé sur les connaissances sémantiques du jugement. Nous avons identifié les indicateurs linguistiques et les vocabulaires réservés du domaine qui signalent les changements thématiques. Pour déterminer les thèmes des segments, les marqueurs les plus importants sont :

- les indicateurs linguistiques qui se trouvent dans les titres des sections
- les positions des paragraphes
- les marqueurs linguistiques et les expressions prototypiques du domaine
- l'ordre d'apparition des segments thématiques dans la progression du jugement

Pour la segmentation, nous avons écrit les règles sémantiques qui contrôlent la présence des marqueurs linguistiques en considérant les contextes dans lesquels les termes sont utilisés (section 6.5).

Nous avons identifié les titres plus significatifs pour les sections qui séparent les différentes étapes du jugement. La table 5.1 montre des exemples de marqueurs identifiés dans les titres et les exemples des titres observés dans les jugements classés selon leur thème dans le document.

Les DONNÉES DE LA DÉCISION peuvent être extraites par les informations présentées dans la première page couverture du jugement. Dans le champ *autorités et doctrines citées*, nous gardons les références et les renvois juridiques cités dans le document. Cette section est séparée du texte de jugement par un titre comme *Reasons for order* ou *Reasons for judgment*. Un exemple des informations classées sous ce thème est illustré par la table 5.2, ce sont les données du jugement analysé présenté dans la table 5.6.

Pour la majorité des jugements des cours canadiennes, le juge commence avec une introduction sur le sujet qui constitue les premiers paragraphes donnant des informa-

<p>Nom de la cour de décision : Cour fédérale du Canada Lieu de l'audience : Ottawa Date : 31/12/97 Numéro de greffe : T-1989-96 Identification des parties : Commissaire aux langues officielles du Canada, Requéran - et - Air Canada, Intimée Intitulé du jugement : Langues officiels du Canada Autorités et doctrines citées : Section 78 de loi sur les langues officiels du Canada</p>
--

TAB. 5.2 – Un exemple de segment de DONNÉES DE LA DÉCISION dans un jugement

tions importantes sur l'affaire, le thème attribué à ce segment est INTRODUCTION. Par exemple le segment INTRODUCTION de la table 5.6, situé au premier paragraphe du jugement présente le problème du service bilingue de la compagnie Air Canada.

Après l'INTRODUCTION, nous trouvons les informations sur les faits et l'histoire du litige, ce segment est associé au thème CONTEXTE. Les titres de la section CONTEXTE peuvent, par exemple, contenir *Facts*, *The factual background* ou *Agreed statement of facts*. CONTEXTE contient deux niveaux de discours : le premier concerne la présentation des parties et leur demande, le deuxième comprend la partie narrative du jugement dans laquelle le juge raconte les événements et les faits. Pour la présentation des parties, les marqueurs sont : *appellant*, *applicant*, *defendant*, *plaintiff*, *respondent* et *intervenor* suivis de leur identification (ex. *the applicant is company X*). Pour la demande, les verbes suivants sont utilisés *advise*, *indicate*, *request*, *claim*, *allege*, *concern*, etc.

Afin d'augmenter la précision de l'identification des parties, nous avons défini certaines règles qui tiennent compte de la sémantique. Par exemple, si la phrase suivante est placée dans le segment CONTEXTE, le patron *the respondent is a* n'est pas une présentation de partie : *On February 14, 1996, the Minister issued, pursuant to subsection 70(5) of the Immigration Act, an opinion that **the respondent is a** "danger to the public".* Les règles sémantiques vérifient donc les points suivants : est-ce que cette phrase reflète une opinion ou un point de vue spécifique ? est-ce que cette phrase reflète une définition dans un cas particulier (*the person constitutes a danger to the public in Canada is ...*) ? est-ce une citation ?

Dans la partie narration du discours, on peut extraire ce qui s'est passé entre les deux parties et l'histoire du problème en litige. Le texte narratif du thème CONTEXTE peut être identifié grâce au temps des verbes exprimés au passé.

Dans certains jugements, plusieurs paragraphes, avec un titre comme *Respondent's*

position, Arguments of party ou *Applicant's submissions*, expliquent la position d'une partie devant le juge. Ce segment SOUMISSION contient les points de vue des avocats des parties et les différents arguments présentés au juge, alors qu'ils peuvent être non pertinents pour arriver à la solution du problème juridique. Normalement ce thème se divise en deux sections afin d'exprimer la position de chaque partie en litige. L'identification de ce thème par le segmenteur est importante, parce qu'il sépare les frontières des autres thèmes. Ce segment isolé est considéré moins pertinent pour le résumé donc il sera éliminé dans le prochain module de filtrage.

La section ISSUES ou questions en litige comprend des questions de fait ou de droit auxquelles le juge doit répondre afin de déterminer le résultat du litige. Cette section a des titres comme *The issues* ou *The questions of law*. Selon le cas, le juge peut avoir à répondre à une ou plusieurs questions, mais il arrive que ces questions ne soient pas expressément posées par le juge dans ses motifs. Ces questions de droit sont importantes parce qu'elles précisent le but du raisonnement du juge, le thème ISSUES de la table 5.6 montre un exemple. Ces deux dernières sections précèdent la partie RAISONNEMENT JURIDIQUE. Nous les avons considérées comme une borne entre deux thèmes CONTEXTE et RAISONNEMENT JURIDIQUE. Après avoir exposé les informations historiques sur le sujet, le juge commence le RAISONNEMENT JURIDIQUE, au point de vue juridique c'est la partie la plus importante dans une décision, qui amène la lecture du CONTEXTE jusqu'à la conclusion finale de la cour. Ce thème explique ce que le juge retient des arguments pertinents des parties et comment il applique le droit sur le cas. Celle-ci permet au juge de trouver une solution pour le problème en répondant aux questions de droit et justifier sa décision. Ce segment peut avoir un titre comme *Analysis* ou *Discussion*. Comme mentionné dans la partie CONTEXTE, dans la plupart des jugements, deux styles différents sont présents dans la décision selon leur position dans le texte. Dans les sections précédant le RAISONNEMENT JURIDIQUE, le juge s'exprime avec un style narratif pour décrire les faits résultant du litige; c'est le juge qui analyse les lois et les précédents judiciaires pour arriver à une conclusion. Donc dans le discours du thème RAISONNEMENT JURIDIQUE le juge s'exprime par *I*. Les exemples des expressions utilisées dans le RAISONNEMENT JURIDIQUE sont : *In reviewing the sections No. of the Act, Pursuant to section No., As I have stated, In the present case, The case at bar is, I am of the view that, etc.*

La dernière section du jugement exprime la CONCLUSION où le juge prononce le **dispositif** dans laquelle est indiqué ce qui est décidé. Habituellement, les phrases contenant le dispositif apparaissent dans les derniers paragraphes du jugement et leur forme est au passif. La présence des marqueurs linguistiques afin de déterminer les résultats du jugement est importante. Quelques exemples de phrases donnant le dis-

positif sont : *The motion is dismissed, the application must be granted*. Nous avons identifié une liste des marqueurs de classe des verbes qui expriment la décision de la jurisprudence : *allow, deny, dismiss, grant, refuse, award, etc.*

Dans un jugement, on peut avoir une section contenant un paragraphe comme l'ordonnance (*Order*), avec une ou deux phrases qui décrivent le résultat très bref prononcé par la cour. Cette section peut être placée au début ou à la fin du jugement après la conclusion.

Nous avons considéré la signature du juge comme la fin du jugement avec les marqueurs : le nom ou initiales du juge, son titre (*Judge*), la date et le lieu de l'audience. Par la suite, le jugement peut être complété avec les informations sur les références complètes de tous les articles de lois cités, les doctrines citées, etc., qui ne sont pas pertinents pour le résumé.

5.2.2 Filtrage

L'objectif du filtrage est d'identifier les segments qui peuvent être supprimés dans les documents sans perdre les informations pertinentes pour le résumé. Dans les cas de citation, comme dans l'exemple de la table 5.3, la présence des marqueurs signale un changement thématique de la citation. Cette méthode permet de réduire la quantité de texte à analyser.

Dans un jugement, les citations occupent un volume important dans le texte soit environ 30% du jugement, alors que leur contenu est moins important pour le résumé, donc nous les considérons comme des exemples négatifs. Pour cette raison, à l'intérieur des segments thématiques, nous identifions les citations à supprimer. Les citations comprennent deux catégories : la première, les soumissions des parties sont les prétentions et arguments des parties concernant les points de vues des parties sur le litige ; la seconde, les citations sont des prétentions en droit concernant les citations des articles de lois applicables sur l'affaire. Nous identifions toutefois les résultats des citations à conserver dans les unités textuelles candidates pour le résumé.

L'identification de citation est basée sur deux types de marqueurs : les marqueurs directs et indirects. La première catégorie de marqueur direct contient les indicateurs linguistiques. La table 5.4 montre les marqueurs linguistiques identifiés en trois classes : verbes de citation, concepts (nom, adverbe, adjectif) et indices complémentaires. Les verbes que nous avons identifiés en anglais, concordent avec les travaux linguistiques sur l'identification des verbes de citation par Mourad [Mourad, 2000] [Mourad, 2001]

<p>Intitulé du jugement : l'accès d'information Numéro de greffe : T-1819-98</p>
<p>In support of its application the plaintiff maintained that the Department had erred in fact and in law when it decided that the plant inspection reports could be disclosed to Mr. D., since the tests for exceptions to the disclosure of documents contained in s. 20(1)(c) and (d) of the Act had not been met. Section 20(1) of the Act reads as follows :</p>
<p>20. (1) Subject to this section, the head of a government institution shall refuse to disclose any record requested under this Act that contains :</p> <ul style="list-style-type: none"> (a) trade secrets of a third party. (b) financial, commercial, scientific or technical information that is confidential information supplied to a government institution by a third party and is treated consistently in a confidential manner by the third party. (c) information the disclosure of which could reasonably be expected to result in material financial loss or gain to, or could reasonably be expected to prejudice the competitive position of, a third party. (d) information the disclosure of which could reasonably be expected to interfere with contractual or other negotiations of a third party.

TAB. 5.3 – Exemple d'une citation : la ligne entre les deux paragraphes montre un changement thématique qui signale la présence de la citation. Les mots en gras sont les marqueurs de citation. Cette unité pointe vers le bloc énuméré, qu'il va aussi considérer comme une citation parce que les phrases de ce bloc sont liées avec des intégrations linéaires à la première unité contenant des marqueurs de citation.

effectuée sur des textes en français, des types des articles des journaux et des rapports scientifiques.

La deuxième catégorie comprend les marqueurs de citation indirects. Les unités textuelles citées indirectement sont les unités voisines des phrases citées directement. Dans la table 5.3, les marqueurs de citation directe sont les mots écrits en gras **Section 20(1) of the Act reads as follows :**, mais les unités textuelles qui suivent cette phrase sont aussi les unités citées. Nous avons donc intégré un mécanisme d'identification des intégrations linéaires entre les phrases suivantes de la première phrase citée avec les marqueurs directs. Afin de déterminer les bornes des segments de citation, il est important de préciser le commencement et la fin des citations. Pour chaque unité marquée comme citation, nous avons considéré un voisinage qui contient les unités textuelles (phrase ou paragraphe) situées avant ou après de l'unité marquée citation, avec la probabilité qu'ils soient une citation. Nous avons développé des règles sémantiques contrôlant le voisinage de l'unité citée et qui seront décrites dans la section 6.6 sur l'implémentation du système.

Classe des verbes	Classe des concepts	Classe des indices
<i>conclude, define, indicate, provide, read, reference, refer, say, state, summarize, ...</i>	<i>following, section, subsection, page, paragraph, pursuant, ...</i>	Ponctuation : deux-points et guillemet ; Prépositions : <i>at (page), (pursuant) to</i> ; Numéros : (section, page, article de loi) ; Subordonnées relatives (conjunction) : <i>that, as, if</i> .

TAB. 5.4 – Les marqueurs linguistiques de citation

5.2.3 Sélection des unités textuelles

La prochaine étape du traitement des segments est la sélection des unités textuelles afin de construire une liste d'unités saillantes candidates pour chaque segment thématique du résumé. Il reste à déterminer quelle information doit absolument se retrouver dans le résumé. Pour construire une liste des phrases candidates saillantes, nous avons utilisé des méthodes linguistiques et statistiques. Comme méthode linguistique, nous avons développé des règles sémantiques basées sur les marqueurs linguistiques. La table 5.5 montre les marqueurs linguistiques utilisés dans ce module de sélection des unités textuelles qui signalent les phrases importantes du jugement. Les marqueurs sont classifiés d'après leurs thèmes dans le jugement. Les marqueurs de l'INTRODUCTION identifient les unités qui signalent le problème et éventuellement les références juridiques utilisées durant le raisonnement. Les marqueurs du CONTEXTE reconnaissent les fragments textuels qui présentent les deux parties et le problème en litige, mais les points de vue des parties ne sont pas pertinents pour le résumé. Les marqueurs du RAISONNEMENT JURIDIQUE déterminent les fragments textuels concernant les analyses du juge et la conclusion retenue par ces analyses. Les marqueurs de la CONCLUSION indiquent les phrases du dispositif qui décrit la décision finale du juge. Les phrases identifiées par les modules linguistiques ont une pondération plus élevée par rapport aux phrases pondérées par les calculs statistiques.

Afin de pondérer les unités, nous avons attribué un poids à chaque phrase en nous basant sur des fonctions heuristiques. La somme des poids permet de calculer un score pour chaque phrase. Pour développer le module de sélection et attribuer les poids, nous avons pris en compte les points suivants :

- La position des phrases dans le paragraphe ; la première phrase du paragraphe et la dernière phrase sont plus importantes.

Thèmes	Marqueurs linguistiques
INTRODUCTION	Concepts : <i>decision, motion, application, determination, order, proceeding, against, rule, reason</i> Verbes : <i>raise, strike, determine, describ, declare, date</i> Expressions : <i>application for judicial review, application to review a decision, motion filed by, Statement of Claim</i>
CONTEXTE	Concepts : Les parties (<i>appellant, applicant, defendant, intervenor, plaintiff, respondent</i>) Verbes : <i>advise, indicate, request, claim, allege, concern</i>
RAISONNEMENT JURIDIQUE	Concepts : <i>opinion, conclusion, summary, because, cost, action, view, reason, I, my</i> Verbes : <i>note, accept, summarise, scrutinize, think, say, satisfy, discuss, conclude, find, believe, reach, indicate, persuade, agree, see, review, recommend, answer</i> Expressions : <i>in the case at bar, for all the above reasons, in my view, my review of, in view of the evidence, in the result, finally, thus, consequently</i>
CONCLUSION	Concepts : <i>motion, application, action, les parties (appellant, applicant, defendant, intervenor, plaintiff)</i> Verbes : <i>allow, deny, dismiss, grant, refuse, award</i> Expressions : <i>for (all) the above reasons, for all of the foregoing reasons</i>

TAB. 5.5 – Les marqueurs linguistiques utilisés dans le module de sélection des unités textuelles qui signalent les phrases importantes dans le jugement

- La position des paragraphes dans le segment thématique : les premier et dernier paragraphes dans le segment ont des valeurs importantes.
- La position des paragraphes dans le document : le premier et les deux derniers paragraphes dans le document sont plus importants.
- La pondération basée sur le calcul statistique des fréquences des mots, pour les termes individuels et les termes dans le contexte du document. Un module calcule les fréquences des mots et un autre calcule $\sum tf * idf$ pour chaque phrase où tf désigne *term frequency* (la fréquence du terme dans le document) et idf *inverted document frequency* mesure si ce terme est discriminant ou non-uniformément distribué dans le corpus. Un terme qui a une valeur de $tf * idf$ élevée doit être à la fois important dans ce document et apparaître rarement dans les autres documents. C'est le cas quand un terme correspond à une caractéristique importante et unique d'un document qui peut exprimer un topique de document. Le $\sum tf * idf$ sur les mots d'une phrase donne le poids de la phrase.
- Les valeurs des marqueurs linguistiques : pour les phrases qui contiennent des marqueurs linguistiques (classe des verbe-concept-indices, expressions prototy-

piques et vocabulaires contrôlés du domaine), nous leur avons affecté une priorité supérieure aux autres critères.

Le résultat de cette étape est une liste de phrases ayant les poids les plus élevés.

5.2.4 Production du résumé

Une fois sélectionnées les unités candidates potentielles pour le résumé, cette étape choisit les unités pour le résumé final et les combine afin de produire un résumé d'environ 10% du jugement. La présentation finale du résumé est sous la forme de fiche de résumé. Cette fiche présente les phrases importantes du jugement en indiquant les thèmes des phrases du résumé. Pour les phrases extraites de chaque thème, selon la structure organisationnelle du jugement, l'utilisateur peut obtenir des informations sur l'importance d'une phrase, sa situation dans le document et son interprétation juridique dans le jugement. Celle-ci augmente la cohérence et la lisibilité des phrases extraites et améliore la qualité générale du résumé. Ce module contrôle les phrases candidates pour le résumé final, il élimine les phrases candidates par module de sélection qui sont annotées comme citations par le module de filtrage.

Le critère de sélection des unités est basé sur la pondération du segment thématique contenant les unités candidates. Selon nos analyses de corpus présentées dans la table 4.3, la distribution de l'information dans les résumés des arrêtistes, mesure l'importance des segments thématiques.

Lors de cette étape de sélection de la liste des unités candidates, nous choisissons les unités du segment thématique INTRODUCTION avec les scores plus élevés jusqu'à concurrence de 10% de la taille de résumé. Dans le segment CONTEXTE, les unités sélectionnées occupent 24% de la longueur du résumé. La contribution du segment RAISONNEMENT JURIDIQUE est de 60% et les unités du thème CONCLUSION occupent 6% du résumé.

5.3 Algorithme

Au prochain chapitre, nous décrivons l'implémentation de LetSum, notre prototype fondé sur notre méthode de résumé automatique pour les textes juridiques. Nous rappelons ici les points les plus importants de notre méthodologie. L'implémentation de notre modèle de conception de la figure 5.1 comprend les étapes suivantes :

1. Segmentation thématique : nous avons utilisé les conditions suivantes afin de découper un segment par thème. La première condition satisfaite arrête le traitement.
 - Présence des titres des sections et classement parmi les catégories des titres (table 5.1) : une section avec un titre significatif va segmenter thématiquement avec le thème identifié par son titre.
 - Position absolue d'un segment : le premier paragraphe est une introduction et les deux derniers paragraphes sont la conclusion.
 - Position relative d'un segment (selon le résultat de notre étude de corpus, les segments thématiques identifiés dans le jugement sont linéairement ordonnés) : si deux sections sont thématiquement identifiées et qu'une section entre ces deux segments n'a pas encore une étiquette de thème, alors la position de cette section est un marqueur.
 - Présence des marqueurs linguistiques : LetSum calcule le nombre de marqueurs de segmentation pour les paragraphes. Si une section a plus de marqueurs de même catégorie, alors cette section va être étiquetée par rapport à la catégorie des marqueurs qu'elle contient.
 - Identification des styles narratif et direct (indicateur de la borne des segments CONTEXTE et RAISONNEMENT JURIDIQUE) : détection de deux styles typographiques grâce aux temps et aspect des groupes verbaux.
2. Filtrage de citations qui comporte quatre étapes :
 - Identification des marqueurs et des patrons d'unités textuelles ;
 - Extraction de la phrase contenant ces unités ;
 - Détection des marqueurs d'intégration linéaire : LetSum identifie d'abord les phrases contenant les unités énumérées, ensuite il regroupe ces phrases dans un bloc. Le système vérifie la condition de relation de citation entre l'unité annotée citation avec le bloc énuméré. Si l'unité de citation pointe vers le bloc énuméré alors ce bloc sera aussi considéré comme une citation.
 - Contrôle de la citation pour les voisinages (avant ou après) des unités citées dans le cas d'absence des marqueurs d'énumération.
3. Sélection des unités saillantes : LetSum calcule un poids pour chaque phrase dans le jugement d'après les fonctions heuristiques expliquées au chapitre 6 basées sur les informations suivantes : la position des paragraphes dans le document, la position des phrases dans le paragraphe, les marqueurs linguistiques, les *cue-phrases*, les vocabulaires contrôlés du domaine juridique. Les modules statistiques du système calculent les fréquences des mots et calculent $\sum tf * idf$ pour chaque phrase.
4. Production de la fiche de résumé : LetSum produit un résumé court en éliminant

les unités moins importantes de la liste des unités candidates au niveau de chaque segment thématique. Le système ne garde que les unités textuelles dans la limite de la taille autorisée du segment, selon les statistiques de la table 4.3. Le système génère une fiche de résumé en assemblant les unités sélectionnées de chaque segment thématique en indiquant le thème des fragments dans la jurisprudence.

5.4 Un exemple d'une fiche de résumé

La table 5.6 montre un exemple des fragments textuels pertinents sélectionnés par LetSum appliqué à un jugement de la Cour fédérale du Canada (2468 mots). Les informations dans le segment de DONNÉES DE LA DÉCISION sont les informations pertinentes extraites de la page de couverture du jugement et les informations dans le champ autorités et doctrines citées sont les traces des références citées dans le jugement. Le module de **Segmentation thématique** sépare le texte en segments thématiques (les thèmes sont écrits à la partie gauche des cases dans la table 5.6). Le module de **Filtrage** supprime la phrase qui signale la citation ainsi que les paragraphes cités énumérés (paragraphe (15) marqués Citation dans la table 5.6). Module de **Sélection** choisit les unités textuelles (montré en gras dans la table 5.6) pour chaque segment thématique. Les unités sont sélectionnées d'après leurs thèmes dans le jugement. À partir de toutes ces informations, le module de **Production** contrôle la taille de résumé, si l'ensemble des phrases sélectionnées dépasse 10% du texte source, il élimine les certaines phrases candidates selon l'importance des segments thématiques.

INTRODUCTION	{	(1) An order was made by this Court on February 4, 1997 authorizing the respondent (Air Canada) to raise preliminary objections to the notice of an originating motion filed by the applicant (the Commissioner). As a result, this motion filed by Air Canada on March 18, 1997 raises six alternative preliminary objections asking the Court to strike out in part the motion made by the Commissioner on September 6, 1996 under section 78 of the Official Languages Act.
CONTEXTE	{	<p>1. Facts</p> <p>(2) The Commissioner's originating motion, which was filed with the consent of the complainant Paul Comeau, concerns Air Canada's failure to provide ground services in the French language at the Halifax airport.</p> <p>(3) The Commissioner's motion is filed by the complainant Paul Comeau.</p> <p>...</p>
ISSUE	{	(14) Do those documents filed by the Commissioner in support of his motion that concern similar complaints and relate to all the ground services provided by Air Canada at the Halifax airport exceed the scope of the remedy provided for in the Act ?
RAISONNEMENT	{	(18) In my view, the purpose of section 79 is to enable the Commissioner to prove to the Court that there is a systemic problem and that it has existed for a number of years. Unless all similar complaints are filed in evidence, the Court cannot assess the scope of the problem and the circumstances of the application.
CONCLUSION	{	<p>7. Conclusion</p> <p>(30) This motion to strike by Air Canada with respect to the preliminary objections must accordingly be dismissed.</p>

TAB. 5.6 – Exemple de l'analyse d'un jugement pour construire une fiche de résumé. Le jugement source est divisé en segments thématiques. Les unités montrées en gras sont jugées pertinentes et les phrases les contenant constitueront le résumé final.

Chapitre 6

Implémentation de LetSum

Dans ce chapitre, nous décrivons LetSum, l'implantation de notre méthode de résumé automatique basée sur les analyses syntaxiques et sémantiques décrites au chapitre précédent. Le système est capable de produire un résumé court sous forme d'une fiche de résumé. Une fiche de résumé produite par LetSum est une table dans laquelle les phrases importantes du document sont inscrites et associées à leur thème.

6.1 Introduction

La production du résumé automatique de LetSum est réalisée en quatre modules (section 5.2) dans l'environnement d'ingénierie linguistique GATE¹ (*General Architecture for Text Engineering*) [Cunningham *et al.*, 2002]. L'entrée du système est un document de jurisprudence qui peut avoir un des formats XML, HTML, SGML, RTF ou un texte sans balise.

Les étapes suivantes sont effectuées :

- Identification du corps du texte du jugement : la séparation du début du jugement comportant les données de la décision et l'identification de la fin comportant les références et la signature du juge,
- Division du texte en sections, paragraphes, phrases, mots, nombres et ponctuations,
- Détection des sections du texte et des titres,
- Segmentation thématique du jugement,
- Identification des citations,

¹<http://gate.ac.uk/>

- Détermination des blocs de citations,
- Extraction des phrases saillantes pour chaque thème,
- Production de la fiche de résumé en transformant la sortie en format XML en une table de résumé en format HTML.

Nous présentons d'abord l'environnement de notre implantation et nous décrivons le développement des modules de LetSum basé sur les connaissances linguistiques juridiques et les méthodes statistiques.

6.2 Architecture de GATE

Cet environnement définit une architecture permettant de construire des applications à partir des modules indépendants. La version 2.2 de GATE fournit des modules permettant d'effectuer les tâches suivantes : séparation des textes en occurrences, segmentation en phrases, étiquetage grammatical, reconnaissance d'entités nommées et résolution de liens de coréférence. Les modules ordonnés au besoin dans un pipeline construisent une application fonctionnelle.

Cet environnement offre une interface graphique de développement qui permet de visualiser les annotations sorties de chaque module et aussi d'annoter manuellement des documents afin de constituer un corpus de référence.

Le développement d'un nouveau module d'annotation ou la modification des modules originaux sont possibles soit en utilisant la bibliothèque de Java soit en langage JAPE (*Java Annotation Patterns Engine*). Un module d'annotation est basé sur la détection de certaines expressions, exécuté en deux étapes.

La première étape est l'identification des marqueurs par un *gazetteer*. Les marqueurs sont stockés comme symboles terminaux dans la grammaire. La deuxième étape du traitement est l'identification de patrons syntaxiques et sémantiques exprimés dans le formalisme JAPE. Un transducteur exécute les patrons basés sur des expressions régulières afin d'annoter le texte. Une grammaire de JAPE comporte un ensemble de phases, chaque phase étant un ensemble de règles sous forme de patron et action. Dans une règle, si les patrons sont satisfaits alors une action (ex. attribution une étiquette d'annotation) pourra être déclenchée. La partie droite d'une règle qui contient des patrons doit être écrite en JAPE mais la partie gauche, ce qui contient les actions, peut être écrite en JAPE ou en JAVA.

Pour l'annotation des unités textuelles, on peut ordonner les règles en définissant

une fonction de priorité. Par défaut, les unités plus longues qui satisfont les conditions auront la priorité, mais il est possible de donner une priorité d'exécution des règles. Une règle plus prioritaire va d'abord être examinée, si toutes ses conditions sont satisfaites alors une étiquette d'annotation sera attribuée aux unités en question et le système arrête l'exécution des autres règles sur la même annotation.

GATE offre aussi la possibilité d'extraire certaines entités nommées (comme les noms des personnes, dates, lieux, etc.) et des coréférences.

6.3 Implémentation

L'implémentation de LetSum, suit le modèle de conception de la figure 5.1 avec les étapes suivantes :

1. **Prétraitement** qui découpe le texte en unités textuelles et identifie les entités nommés nécessaires.
2. **Segmentation thématique** qui détecte la structure de document et les thèmes.
3. **Filtrage** qui annote les segments à supprimer pour le résumé.
4. **Sélection des unités saillantes** qui repère une liste des phrases candidates les plus pertinentes du texte.
5. **Production de la fiche de résumé** qui supprime les certains candidates de module de sélection annotés comme les citations par le module de filtrage. Il contrôle aussi la taille du résumé pour les présenter dans une fiche de résumé.

Dans ce chapitre à partir de la table 6.2 à table 6.14, pour faciliter la lecture, nous présentons conceptuellement les règles sémantiques que nous avons implémentées en JAPE et JAVA, en utilisant une notation inspirée de la BNF. Cette notation contient les caractères suivants :

- Un mot écrit en *italique* présente un terminal.
- Un terminal qui commence par une majuscule montre qu'il faut que ce mot apparaisse au début de la phrase.
- La flèche (\rightarrow) signifie un patron de condition/action. La partie gauche présente les conditions du patron syntaxique. Si le patron concorde avec un segment du texte alors une étiquette du type indiqué dans la partie droite sera attribuée à ce segment.
- Les parenthèses () regroupent un bloc des mots dans le patron.
- L'alternative (|) montre une unité qui peut être choisie parmi plusieurs.
- Les crochets ([]) englobent une partie optionnelle.

- Un mot tout en MAJUSCULE est une étiquette d’annotation.
- Les guillemets (” ”) encadrent une entité qui doit apparaître telle quelle dans le patron.
- Les accolades suivies de plus ({})+ englobent des entités qui peuvent être répétées plus d’une fois.

Dans les prochaines sections nous présentons les différents composants de LetSum à travers les règles sémantiques.

6.4 Prétraitement

L’entrée de LetSum est un jugement en anglais en format HTML. Afin de traiter le document source dans la phase de prétraitement, le système sépare les unités textuelles à analyser. Pour cette tâche, l’exécution des sept étapes est nécessaire : dans les quatre premières étapes nous avons utilisé les outils de GATE alors que nous avons programmé les trois derniers.

1. *Tokenizer* sépare le document en occurrences et donne à chacun un numéro d’identification en distinguant les mots écrits en majuscules, minuscules, tout en majuscules ou mixte.
2. *Sentence splitter* isole des phrases en fonction de ponctuations comme le signe de fin de phrase. Devant un point, il vérifie le mot avec une liste d’abréviations. Les abréviations peuvent être un titre (Mr.), une initiale (J. Baker), un nom de compagnie, etc. Nous avons ajouté les abréviations courantes du domaine juridique à cette liste. Notre référence pour ces abréviations est la liste fournie par la bibliothèque de droit de l’Université de Montréal, diffusée à <http://www.bib.umontreal.ca/DR/abrev.htm>.
3. *Part-of-Speech tagger* ajoute une catégorie syntaxique à chaque mot selon la technique de Hepple [Hepple, 2000]. La table 6.1 donne un exemple de l’analyse d’une phrase par ce taggeur.
4. *ANNIE*² (*A Nearly-New Information Extraction system*) de GATE retrouve les entités nommés de nom de personne, de date et de lieu.
5. Séparation du texte en paragraphes. pour un document HTML, on utilise la balise <P> montrant le début et la fin de paragraphe. Pour les document non HTML, la segmentation en paragraphes se fait en utilisant les caractéristiques suivantes :

²<http://gate.ac.uk/ie/annie.html>

<p>Phrase : <i>For all of the foregoing reasons, this application will be dismissed.</i></p> <p>Analyse syntaxique : <i>For (IN) all (DT) of (IN) the (DT) foregoing (VBG) reasons (NNS), this (DT) application (NN) will (MD) be (VB) dismissed (VBN).</i></p>
<p>Etiquettes de <i>Part-of-Speech</i> utilisées dans taggeur de Hepple</p> <p>NN - noun - singular or mass</p> <p>NNS - noun - plural</p> <p>VB - verb - base form</p> <p>VBG - verb - gerund or present participle</p> <p>VBN - verb - past participle</p> <p>DT - determiner</p> <p>IN - preposition or subordinating conjunction</p> <p>MD - modal</p>

TAB. 6.1 – Exemple d’une phrase analysée par l’analyseur syntaxique

un paragraphe commence par un mot majuscule qui suit un titre, une ligne vide ou un marqueur d’énumération.

- Séparation du texte en sections en identifiant les titres des sections. Après avoir identifié les titres des sections, LetSum classe les titres significatifs en fonction des marqueurs linguistiques expliqués à la section 5.2.1. La table 6.2 montre les règles pour catégoriser des titres des sections en six thèmes : INTRODUCTION, CONTEXTE, ISSUE, SOUMISSION, RAISONNEMENT, CONCLUSION. Nous détectons aussi les deux titres de l’ordonnance et la dépense. Par exemple pour déterminer un titre INTRODUCTION il faut un indicateur comme *introduction* ou *summary* dans le titre possiblement précédé d’un numéro de la section.
- LetSum identifie le début et la fin du texte du jugement, que nous appelons le corps du jugement. La table 6.3 montre que pour séparer les DONNÉES DE LA DÉCISION, comme le nom de la cour, le lieu, la date, le numéro du dossier, les parties et l’intimée du jugement (section 5.2.1) du corps du jugement, il faut un titre contenant l’expression suivante : *reasons for decision* ou *judgment* ou *order* suivis par une ordonnance (*and order*) ou une dépense (*pursuant to cost*).

Le juge met habituellement son nom ou son initiale avant le commencement du texte. Il peut arriver que le jugement commence avec le titre *prothonotary*, protonotaire qui est le fonctionnaire attaché à la juridiction et qui peut exercer certains pouvoirs judiciaires.

La fin de la décision est la frontière de la séparation du corps du jugement avec les informations complémentaires comme les références complètes. Les indicateurs de la fin du jugement sont :

TITRE_INTRODUCTION	→	[NUMÉRO_SECTION] <i>introduction</i> <i>summary</i>
TITRE_CONTEXTE	→	[NUMÉRO_SECTION] <i>background</i> <i>facts</i> <i>factual background</i>
TITRE_ISSUE	→	[NUMÉRO_SECTION] <i>issue[s]</i> <i>question[s] of law</i>
TITRE_SOUMISSION	→	[NUMÉRO_SECTION] <i>submissions</i> <i>position</i> <i>argument</i>
TITRE_RAISONNEMENT	→	[NUMÉRO_SECTION] <i>analysis</i> <i>decision</i> <i>discussion</i> <i>jurisdiction</i>
TITRE_CONCLUSION	→	[NUMÉRO_SECTION] <i>conclusion</i> <i>cost</i> <i>disposition</i>
TITRE_ORDONNANCE	→	[NUMÉRO_SECTION] <i>order</i>
TITRE_DÉPENSE	→	[NUMÉRO_SECTION] <i>costs</i>

TAB. 6.2 – Identification et classification des titres significatifs par thèmes

- La signature du juge comprenant le mot *signed* avec le nom du juge et son titre, par exemple *J.F.C.C (judge of federal court of Canada)*.
- Phrase de confirmation du juge pour valider le jugement.
- Certification de la traduction du jugement. Au Canada un jugement peut être écrit en anglais ou français et avoir une traduction officielle, dans ce cas la traduction peut porter la phrase *Certified true translation* avec le nom du traducteur à la fin du texte.

6.5 Segmentation thématique

L'identification des thèmes est fondée sur le plan organisationnel du jugement tel qu'il a été décrit en section 4.1.2. L'ordre des segments thématiques et leur position sont des facteurs importants. Un jugement commence par DONNÉES DE LA DÉCISION avec ses champs. Après avoir exposé ces données, le corps du jugement commence, selon l'ordre suivant : INTRODUCTION, CONTEXTE, ISSUE, SOUMISSION, RAISONNEMENT et CONCLUSION. Le thème ISSUE peut venir avant CONTEXTE ou après SOUMISSION.

Pour l'implantation du module de segmentation thématique de LetSum, nous avons écrit deux types de règles sémantiques. Le premier comprend des règles qui sont basées essentiellement sur la présence des titres significatifs, la position des segments dans le

DÉBUT_JUGEMENT	→	<i>reasons for (decision judgment order)</i> [<i>and order pursuant to cost</i>] (NOM_JUGE NOM_PROTHONOTARY)
FIN_JUGEMENT	→	SIGNATURE DATE VILLE PROVINCE [CONFIRMATION TRADUCTION]
DATE	→	combinaison de jour, mois et année
PROVINCE	→	<i>Ontario</i> <i>ON</i> autre province du Canada
VILLE	→	<i>Ottawa</i> autre ville canadienne
NOM_JUGE	→	NOM_PROPRE INITIALE <i>judge</i> <i>J.</i>
NOM_PROTHONOTARY	→	NOM_PROPRE INITIALE <i>prothonotary</i>
NOM_PROPRE	→	prénom et nom de personne
INITIALE	→	(<i>A – Z</i>)["."] (<i>A – Z</i>)["."]
SIGNATURE	→	<i>Signed</i> <i>Sgd.</i> <i>Sg.</i> <i>S.</i> NOM_PROPRE INITIALE <i>judge</i> <i>J.</i> <i>J.F.C.C.</i> <i>A.C.J.</i> <i>prothonotary</i>
CONFIRMATION	→	<i>I agree</i>
TRADUCTION	→	<i>Certified true translation</i>

TAB. 6.3 – Identification du corps du jugement

document (position absolue) et la position des segments thématiques d’après leur ordre dans la structure de jurisprudence (position relative).

Le deuxième type contient des règles basées sur les marqueurs linguistiques, les expressions prototypiques et les vocabulaires du domaine juridique.

Nous présentons les règles en deux parties dans les sections 6.5.1 à 6.5.6 :

- **Conditions** cherchent à trouver des bornes de début et de fin d’un segment thématique afin d’isoler des paragraphes associés à un thème.
- **Action** déclenchée si les conditions sont satisfaites. Pour les règles de segmentation, l’action consiste à attribuer une étiquette de thème au segment contenant un ou plusieurs paragraphes.

6.5.1 Identification du thème INTRODUCTION

Le thème INTRODUCTION peut être identifié grâce aux trois règles suivantes que nous les avons présenté dans la table 6.4 :

- R1** — Cette règle annote le thème INTRODUCTION en utilisant les titres significatifs. Si un titre TITRE_INTRODUCTION et un TITRE_CONTEXTE existent, alors les paragraphes entre ces deux limites vont être annotés avec le thème INTRODUCTION. Cette règle a une priorité supérieure à celle des autres règles.
- R2** — Si le TITRE_INTRODUCTION est absent, le commencement du jugement sera le DÉBUT_JUGEMENT (table 6.3) et pourra être suivi par le NOM_JUGE. Les paragraphes entre DÉBUT_JUGEMENT et le TITRE_CONTEXTE sont du thème INTRODUCTION.
- R3** — Si aucun titre significatif n'existe, le premier paragraphe du début du jugement sera considéré comme INTRODUCTION.

<p>R1- Présence des titres :</p> <p>Conditions :</p> <p style="padding-left: 40px;">début du segment : TITRE_INTRODUCTION</p> <p style="padding-left: 40px;">fin du segment : TITRE_CONTEXTE</p> <p>R2- Dans absence du titre du début du segment</p> <p>Conditions :</p> <p style="padding-left: 40px;">début du segment : DÉBUT_JUGEMENT [NOM_JUGE]</p> <p style="padding-left: 40px;">fin du segment : TITRE_CONTEXTE</p> <p>R3- Dans absence du titre du segment la position du segment est utilisé</p> <p>Conditions :</p> <p style="padding-left: 40px;">début du segment : DÉBUT_JUGEMENT [NOM_JUGE]</p> <p style="padding-left: 40px;">fin du segment : fin de premier paragraphe</p> <p>Action : Attribution de l'annotation THÈME INTRODUCTION aux unités textuelles entre les deux conditions</p>
--

TAB. 6.4 – Règles pour la segmentation thématique : thème INTRODUCTION

6.5.2 Identification du thème ISSUE

Dans le plan du jugement, ISSUE, ou les questions de droit, se situe après le thème CONTEXTE. Mais dans l'implémentation du module de segmentation, on identifie d'abord le thème ISSUE parce que sa taille ne dépasse pas un paragraphe et ce segment, s'il existe, peut être distingué avec deux marqueurs (table 6.5) : un TITRE_ISSUE ou le marqueur suivant situé au début du paragraphe : *The issue[s] is/are*.

R1- Présence des titres :

Conditions :

début du segment : TITRE_ISSUE

R2- Présence de marqueur :

Conditions :

début du segment : paragraphe commence par *the issue[s]*

Action : Attribution de l'annotation THÈME ISSUE au unité textuelle qui suit le titre

TAB. 6.5 – Règle pour la segmentation thématique : thème ISSUE

6.5.3 Identification du thème CONTEXTE

La reconnaissance du thème CONTEXTE est basée sur les trois règles suivantes (table 6.6) :

- R1** — Le segment commence avec un TITRE_CONTEXTE ou il suit le segment du thème INTRODUCTION. La fin du segment peut être retrouvée à l'aide d'un TITRE_SOUMISSION des parties.
- R2** — Pour la fin du segment en l'absence de TITRE_SOUMISSION des parties, on cherche le thème ou le titre de ISSUE.
- R3** — En l'absence du thème ISSUE et SOUMISSION des parties, le TITRE_RAISONNEMENT est considéré comme la fin du segment.

<p>R1- Conditions :</p> <p style="padding-left: 40px;">début du segment : TITRE_CONTEXTE THÈME INTRODUCTION fin du segment : TITRE_SOUMISSION</p> <p>R2- Conditions :</p> <p style="padding-left: 40px;">début du segment : TITRE_CONTEXTE THÈME INTRODUCTION fin du segment : TITRE_ISSUE THÈME ISSUE</p> <p>R3- Conditions :</p> <p style="padding-left: 40px;">début du segment : TITRE_CONTEXTE THÈME INTRODUCTION fin du segment : TITRE_RAISONNEMENT</p> <p>Action : Attribution de l'annotation THÈME CONTEXTE aux unités textuelles entre les deux conditions</p>

TAB. 6.6 – Règles de segmentation thématique : thème CONTEXTE

6.5.4 Identification du thème SOUMISSION

Dans la segmentation thématique, on identifie le thème SOUMISSION pour ordonner les segments dans le plan du document mais son contenu est moins important pour le résumé. Ce thème va donc être supprimé dans la phase de filtrage. Le thème SOUMISSION peut être identifié avec deux règles (table 6.7) :

R1 — Le segment commence par un TITRE_SOUMISSION et il se termine avec ISSUE.

R2 — Si le thème ISSUE n'existe pas, il va être limité au TITRE_RAISONNEMENT.

<p>R1- Conditions : début du segment : TITRE_SOUMISSION fin du segment : TITRE_ISSUE THÈME ISSUE</p> <p>R2- Conditions : début du segment : TITRE_SOUMISSION fin du segment : TITRE_RAISONNEMENT</p> <p>Action : Attribution de l'annotation THÈME SOUMISSION aux unités textuelles entre les deux conditions</p>
--

TAB. 6.7 – Règles de segmentation thématique : thème SOUMISSION

6.5.5 Identification du thème CONCLUSION

Nous cherchons le thème CONCLUSION avant le thème RAISONNEMENT parce qu'il va être considéré comme borne finale de ce segment. La CONCLUSION est toujours située à la fin du jugement.

La CONCLUSION peut être retrouvée avec trois règles (la table 6.8) :

- R1** — Le segment commence avec un TITRE_CONCLUSION. Les trois indicateurs suivants peuvent terminer ce segment : La signature du juge (FIN_JUGEMENT), TITRE_ORDONNANCE ou TITRE_DÉPENSE.
- R2** — Les paragraphes contenant les quatre expressions suivantes sont considérées comme la conclusion : *for all the above reasons*, *for the foregoing reasons*, *for all of the foregoing reasons*, *in the result*.
- R3** — Si les indicateurs du début de ce thème ne sont pas retrouvés alors selon la position de ce segment, on considère les deux derniers paragraphes comme la CONCLUSION du jugement.

R1- Présence des titres

Conditions :

début du segment : TITRE_CONCLUSION

fin du segment : FIN_JUGEMENT | TITRE_ORDER | TITRE_DÉPENSE

R2- Présence des marqueurs

Conditions :

début du segment : paragraphe commençant par

(*for all the above reasons* |
for [all of] the foregoing reasons |
in the result)

R3- Prendre les deux derniers paragraphes du jugement

Condition :

fin du segment : FIN_JUGEMENT | TITRE_ORDONNANCE | TITRE_DÉPENSE

Action : Attribution de l'annotation THÈME CONCLUSION aux unités textuelles entre les deux conditions

TAB. 6.8 – Règles de segmentation thématique : thème CONCLUSION

6.5.6 Identification du thème RAISONNEMENT

Ce thème est le dernier thème que LetSum identifie. Parce que ce segment n'a pas toujours un titre précis. Il faut retrouver ce thème à l'aide de sa position relative dans le plan du document. Ce segment se situe après les thèmes CONTEXTE, ISSUE et SOUMISSION et avant le thème CONCLUSION (table 6.9).

<p>Conditions :</p> <p>début du segment : TITRE_RAISONNEMENT THÈME ISSUE THÈME SOUMISSION THÈME CONTEXTE</p> <p>fin du segment : TITRE_CONCLUSION THÈME CONCLUSION</p> <p>Action : Attribution de l'annotation THÈME RAISONNEMENT aux unités textuelles entre les deux conditions</p>

TAB. 6.9 – Règle pour la segmentation thématique : thème RAISONNEMENT

À l'aide des règles de premier type de segmentation thématique, nous arrivons à identifier les thèmes dans plus de 60% des cas (voir section 7.4). Pour les thèmes non identifiés, LetSum les complète avec des règles de deuxième type, basées sur les indicateurs linguistiques et les vocabulaires du domaine juridique.

Nous présentons les règles du deuxième type en utilisant la situation suivante : les règles sont regroupées selon un thème précis et l'unité textuelle qui va être annotée est la phrase. Les phrases contenant les indicateurs thématiques pour ce thème sont importantes. Donc, à l'étape de sélection des unités saillantes par LetSum, ces phrases ont des priorités supérieures à celles des autres phrases candidates pour le résumé. D'après ces règles, les paragraphes entre la première phrase contenant des marqueurs du thème et la dernière phrase de ce thème (ou la première phrase du thème suivant) recevront une annotation de segmentation thématique. Nous décrivons les indicateurs linguistiques signalant les phrases pertinentes pour un thème dans la section 6.7 du module de sélection.

6.6 Filtrage

Le module de filtrage identifie les unités moins importantes dans le jugement afin de les supprimer. L'exécution de ce module augmente la précision d'extraction des phrases pertinentes pour les prochains modules de LetSum.

Dans ce module, deux types d'information vont être filtrés, d'abord le segment thématique de *Soumission* des parties, identifié par le segmenteur thématique, ensuite les citations des droit mentionnés dans le jugement.

6.6.1 Détection des citations

Afin de détecter les citations, chaque règle cherche des marqueurs signalant la citation. La table 6.10 montre les huit règles permettant de retrouver les citations. Les six premières règles identifient une phrase de citation, terminée par un deux-points, qui annonce les paragraphes, parfois énumérés comprenant les éléments des articles de lois et les droits applicables au cas. Les deux dernières règles comprennent les deux expressions à la fin des paragraphes ou les phrases cités. Par exemple [*paras. 12*] qui peut être suivi par ... réfère à un numéro de paragraphe dans l'article, le reste du texte étant supprimé par le juge.

Pour l'identification des citations indirectes, qui ne contiennent pas les marqueurs de citation, d'abord LetSum cherche les marqueurs d'intégration linéaire qui enchaînent les éléments cités avec des marqueurs d'énumération. Les marqueurs d'énumération dans les jugements ont les formats suivants :

- (*a*) ou *a*) ou *a*.
- *20.(9)* ou *20.12(9)*
- [*paras. 14*]
- *34(2)(a)* ou *34(a)*
- *i.* ou *XII.*

En l'absence d'énumération, LetSum utilise les heuristiques suivantes afin de contrôler les voisinages des paragraphes cités :

- Un paragraphe suivant une phrase citée (déterminée par les règles 1 à 6 de citation) est une citation.
- Un paragraphe entre deux paragraphes cités est une citation.

Dans les jugements des cours du Canada les juges citent souvent des paragraphes dans la langue originale qui peut différer de la langue du jugement. LetSum détermine les citations en anglais et en français.

R1	→	CITATION_VERBE [ÀNUMÉRO] <i>follow</i> [PAGE] :
R2	→	<i>rule</i> NOMBRE CITATION_VERBE :
R3	→	CITATION_VERBE (<i>as</i> <i>if</i> <i>that</i>) :
R4	→	SECTION NOMBRE CITATION_VERBE :
R5	→	<i>pursuant to</i> (<i>section</i> <i>subsection</i>) NOMBRE <i>of act</i> :
R6	→	<i>which</i> CITATION_VERBE <i>as follows</i> :
R7	→	”[” <i>paras.</i> NOMBRE ”]” ...
R8	→	”[” <i>Citation omitted</i> ”]” ...
<hr/>		
CITATION_VERBE	→	<i>conclude</i> <i>define</i> <i>indicate</i> <i>makereference</i> <i>provide</i> <i>read</i> <i>refer</i> <i>say</i> <i>setout</i> <i>state</i> <i>summarize</i>
ÀNUMÉRO	→	<i>at</i> NOMBRE (<i>and</i> <i>to</i>) NOMBRE
NOMBRE	→	{(0 – 9)[”.”]} +
PAGE	→	<i>page</i> [s] <i>paragraph</i> [s] <i>p.</i> <i>para</i> [s].

TAB. 6.10 – Grammaire d’identification des unités de citation

6.7 Sélection des unités saillantes

Le module de sélection des unités saillantes de LetSum comprend deux méthodes. La première est basée sur des indicateurs juridiques et linguistiques, la deuxième méthode est statistique.

Les règles linguistiques élaborées manuellement sont basées sur nos analyses de comparaison des jugements avec des résumés rédigés par les arrêtiistes.

Pour chaque thème dans le jugement, nous avons identifié les marqueurs particuliers pour un thème précis et indiquant des phrases pertinentes dans le document. Ceci met en évidence l’importance relative d’une phrase par rapport à son thème dans le jugement.

Nous présentons les règles de sélection des unités candidates, classifiées par thèmes, ainsi que les marqueurs linguistiques, catégorisés en classes des verbes, concepts et expressions prototypiques (*cue-phrases*).

R1	→	<i>judicial review [of] a decision [of]</i>
R2	→	MOTION <i>by</i> (MARQUEUR_PARTIE NOM_PROPRES) <i>for</i>
R3	→	MOTION <i>is for</i>
R4	→	PLAINTIFF <i>filed an application</i> <i>to review a decision of</i>
R5	→	(PLAINTIFF MOTION) <i>seek[s]</i>
R6	→	prendre la première phrase de ce thème
<hr/>		
PLAINTIFF	→	NOM_PROPRES <i>plaintiff</i> <i>applicant</i>
MARQUEUR_PARTIE	→	<i>appellant</i> <i>applicant</i> <i>defendant</i> <i>intervenor</i> <i>plaintiff</i> <i>respondent</i>
NOM_PROPRES	→	prénom et nom de personne
MOTION	→	<i>action</i> <i>application</i> <i>claim</i> <i>decision</i> <i>motion</i>

TAB. 6.11 – Grammaire d’identification des phrases saillantes de INTRODUCTION

La table 6.11 présente quatre règles de sélection des phrases du thème INTRODUCTION. Ce thème comprend des paragraphes au début d’un jugement, présentant une information générale sur ses origines, son contenu ou sa structure.

- R1** — repère les phrases de INTRODUCTION qui présente le sujet. Par exemple *The Applicant seeks judicial review of a decision rendered against the applicant.*
- R2** — présente la demande et la partie avec son nom propre ou son titre dans le jugement.
- R3** — mentionne la demande. Par exemple *This motion is for the preliminary determination of a question of law.*
- R4** — montre le sujet du litige et les parties.
- R5** — donne un survol de la déclaration d’une demande, *The applicant seeks an order declaring [...].*
- R6** — En l’absence des marqueurs linguistiques selon la position des phrases, prend la première phrase du segment thématique.

R1	→	<i>filed a statement of claim</i>
R2	→	PLAINTIFF <i>amended a statement of claim</i>
R3	→	MARQUEUR_PARTIE VERBE_ÊTRE
R4	→	[MARQUEUR_PARTIE ' s] MOTION CONTEXTE_VERBE
R5	→	[MARQUEUR_PARTIE ' s] MOTION VERBE_ÊTRE <i>supported by</i>
R6	→	PLAINTIFF CONTEXTE_VERBE
R7	→	phrase commence avec une DATE précise
<hr/>		
PLAINTIFF	→	NOM_PROPRE <i>plaintiff</i> <i>applicant</i>
NOM_PROPRE	→	prénom et nom de personne
MARQUEUR_PARTIE	→	<i>appellant</i> <i>applicant</i> <i>defendant</i> <i>intervenor</i> <i>plaintiff</i> <i>respondent</i>
CONTEXTE_VERBE	→	<i>advise</i> <i>allege</i> <i>claim</i> <i>concern</i> <i>indicate</i> <i>request</i>
MOTION	→	<i>action</i> <i>application</i> <i>claim</i> <i>decision</i> <i>motion</i>
DATE	→	jour, mois et année
VERBE_ÊTRE	→	<i>is</i> <i>are</i> <i>was</i> <i>were</i>

TAB. 6.12 – Grammaire d'identification des phrases saillantes de CONTEXTE

La table 6.12 montre les règles de repérage des phrases du thème CONTEXTE. Ce thème est considéré comme un ensemble d'informations précisant les parties dans le procès et les caractéristiques des événements ainsi que des faits.

R1 — présente les prétentions.

R2 — montre une modification sur la demande par le plaignant, *plaintiff has amended his statement of claim*.

R3 — présente une partie décrite avec le verbe être, *Mr. Masunda is a citizen of Zimbabwe who claimed status as a Convention refugee based on his political opinion*.

R4 — décrit la nature de la demande.

R5 — présente le souhait de la demande en précisant la partie.

R6 — précise le désir du demandeur, *Mr. Barrera requested a humanitarian and compassionate exemption in respect of his application for permanent residence in Canada*.

R7 — identifie les faits et les événements selon leurs dates.

R1	→	<i>I</i> RAISONNEMENT_VERBE
R2	→	<i>In my [respectful]</i> RAISONNEMENT_CONCEPT
R3	→	<i>my</i> RAISONNEMENT_CONCEPT TOBE
R4	→	<i>consequently finally thus</i>
R5	→	<i>(have has)</i> TEMPS <i>to</i>
R6	→	CUE-PHRASES
RAISONNEMENT_VERBE	→	<i>accept agree believe conclude discuss fine indicate note persuade reach review satisfy say scrutinize summarise think</i>
RAISONNEMENT_CONCEPT	→	<i>analysis conclusion decision finding mind opinion review summary view</i>
CUE-PHRASES	→	<i>Most important to note, For these reasons, As a result, I am satisfied Because of I take this to mean that In view of the evidence</i>
TEMPS	→	ANNÉE MOIS JOUR
TOBE	→	<i>is are</i>

TAB. 6.13 – Grammaire d’identification des phrases saillantes de RAISONNEMENT

La table 6.13 représente les règles de sélection des phrases du thème RAISONNEMENT. Les phrases de ce thème sont en forme discursive consistant à mettre en propositions et à résoudre le problème. Les phrases signalant l’opinion du juge et le résultat de son analyse sont importantes.

R1 — présente l’analyse du juge grâce à une liste des verbes que nous avons identifiés.

R2 — présente l’opinion du juge sur le litige.

R3 — montre le motif du jugement.

R4 — présente les phrases conclusives sur le raisonnement du juge.

R5 — identifie les unités présentant le délai donné par le juge pour les dépenses.

R6 — comprend les expressions prototypiques de ce thème.

R1	→	MARQUEUR_PARTIE ' s MOTION [for judicial review] (TOBE (MODALE be)) CONCLUSION_VERBE [with costs]
R2	→	I CONCLUSION_VERBE MOTION
R3	→	MOTION (TOBE (MODALE be)) CONCLUSION_VERBE [with costs]
R4	→	in conclusion
<hr/>		
CONCLUSION_VERBE	→	award allow deny dismiss grant refuse strike
MARQUEUR_PARTIE	→	appellant applicant defendant plaintiff respondent
MOTION	→	action application claim decision motion
TOBE	→	is are
MODALE	→	can could will would shall should must may might

TAB. 6.14 – Grammaire d'identification des phrases saillantes de CONCLUSION

La table 6.14 montre les règles de sélection des phrases du thème CONCLUSION qu'identifie le dispositif. Les phrases relevantes de ce thème donnent des résultats de la décision de la cour et justifient toutes les thèmes précédentes. Les marqueurs linguistiques suivants peuvent être précédés par l'expression : *for all the above reasons* ou *in the result*.

R1 — présente la décision sur la demande du plaignant en passif, *Plaintiff's claim is struck with costs*.

R2 — montre le résultat prononcé par le juge directement.

R3 — mention la décision qui peut comprend le dépense, *this application for judicial review will be dismissed*.

R4 — présente la conclusion finale de la cour.

6.7.1 Module de calculs statistiques

L'implémentation de la deuxième méthode de sélection est fondée sur les calculs statistiques. LetSum calcule un poids pour chacune des phrases du document ; une phrase ayant un poids plus élevé est plus importante. Le poids final de chaque phrase est la somme de trois poids calculés par la fréquence des mots, par la position de phrase dans le paragraphe et par la position dans le segment thématique.

Le calcul de fréquence pour une phrase est basé sur le $tf * idf$ pour chaque mot du document. LetSum calcule la fréquence des mots dans le document pour la valeur de tf . Un mot plus fréquent dans le document est un mot plus important. Mais pour calculer les fréquences il faut d'abord normaliser les mots par la lemmatisation et l'élimination des mots vides. Par exemple, le verbe *dismissed* et *dismiss* sont considérés identiques pour le calcul de la fréquence. LetSum compare ensuite le lemme d'un mot avec une liste de 350 mots vides, les mots sans valeurs sémantiques importantes (prépositions, déterminants, ...). Si un mot appartient à cette liste alors on l'ignore.

Pour le calcul de idf , on regarde la fréquence inversée d'un mot par rapport au corpus. Un mot moins fréquent, avec un idf plus élevé dans le corpus est un mot plus spécifique pour le document. Nous avons calculé idf (*inverted document frequency*) en utilisant la collection des jugements de la Cour fédérale du Canada, qui contient les décisions rendues entre 1992 et 2004. La collection comporte 10317 jugements en anglais disponibles sur la page web de CanLII³. Pour une phrase la somme des valeurs $tf * idf$ pour chaque mot de la phrase donne un poids pour la phrase. On conserve les phrases avec un poids plus grand.

La position des phrases est un autre élément permettant d'augmenter le poids d'une phrase. Selon nos observations sur le corpus, la plupart des phrases extraites dans les résumés humains sont situées au début ou à la fin du paragraphe, il en est de même pour la position d'un paragraphe dans le segment thématique. Le premier et le dernier paragraphe de chaque thème sont plus importants. La somme des poids calculés pour chaque phrase, déterminé par ces trois critères donne le score final d'une phrase.

Pour la sélection des phrases candidates, celles qui sont repérées par les règles linguistiques sont prioritaires. Les phrases sélectionnées par les modules statistiques complètent le résumé jusqu'à environ 10% du texte source.

³<http://www.canlii.org/ca/cas/fct/>

6.8 Production de la fiche de résumé

LetSum produit un résumé en éliminant les unités moins importantes de la liste des unités candidates au niveau de chaque segment thématique. Le système ne garde que les unités textuelles dans la limite de la taille autorisée du segment, selon les statistiques de la table 4.3. Ce module retire ensuite les unités annotées comme citations par le module de filtrage. Le système génère une fiche de résumé en assemblant les unités sélectionnées de chaque segment thématique en indiquant le thème des fragments dans la jurisprudence.

L'implémentation de ce module est effectuée à l'aide d'une feuille de style XSLT traitant la sortie de XML de GATE. Il s'assure que les phrases de résumé final ne sont pas des citations et il contrôle la taille des extraits. Finalement, il affiche les phrases saillantes en les reliant avec leurs thèmes dans une fiche de résumé.

6.9 Sorties de LetSum

Nous présentons un exemple de résumé produit pour un jugement de la Cour fédérale du Canada par LetSum. Les figures montrent les étapes de génération du résumé pour les quatre modules principaux du système : segmentation thématique (section 6.5), filtrage (section 6.6), sélection des phrases (section 6.7) et production de la fiche de résumé (section 6.8).

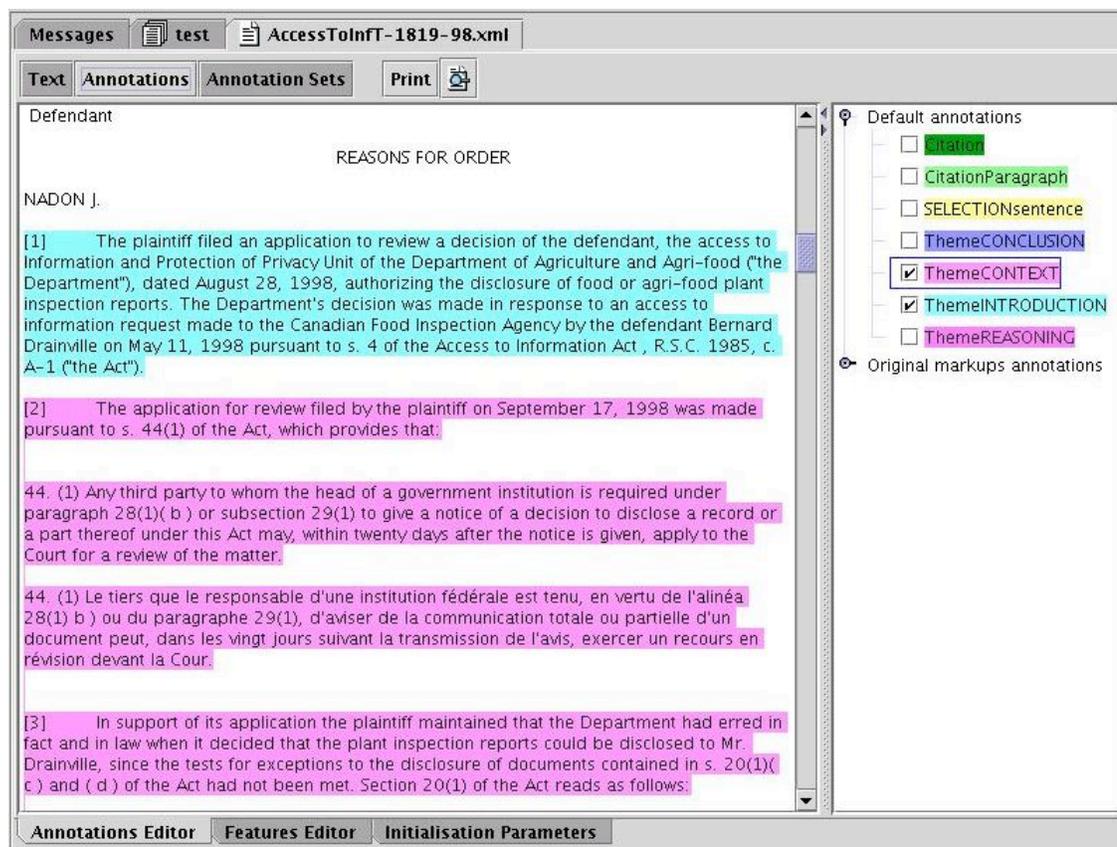


FIG. 6.1 – Sortie de module de segmentation thématique identifiant les thèmes du jugement, sous l'interface de GATE

La figure 6.1 montre un exemple d'identification des segments thématique par LetSum. Ce module est expliqué à la section 6.5. L'interface de GATE permet de visualiser les étapes de production du résumé. Chaque thème est identifié avec des couleurs différentes. L'interface peut également donner accès à la règle de grammaire ayant identifié ce thème.

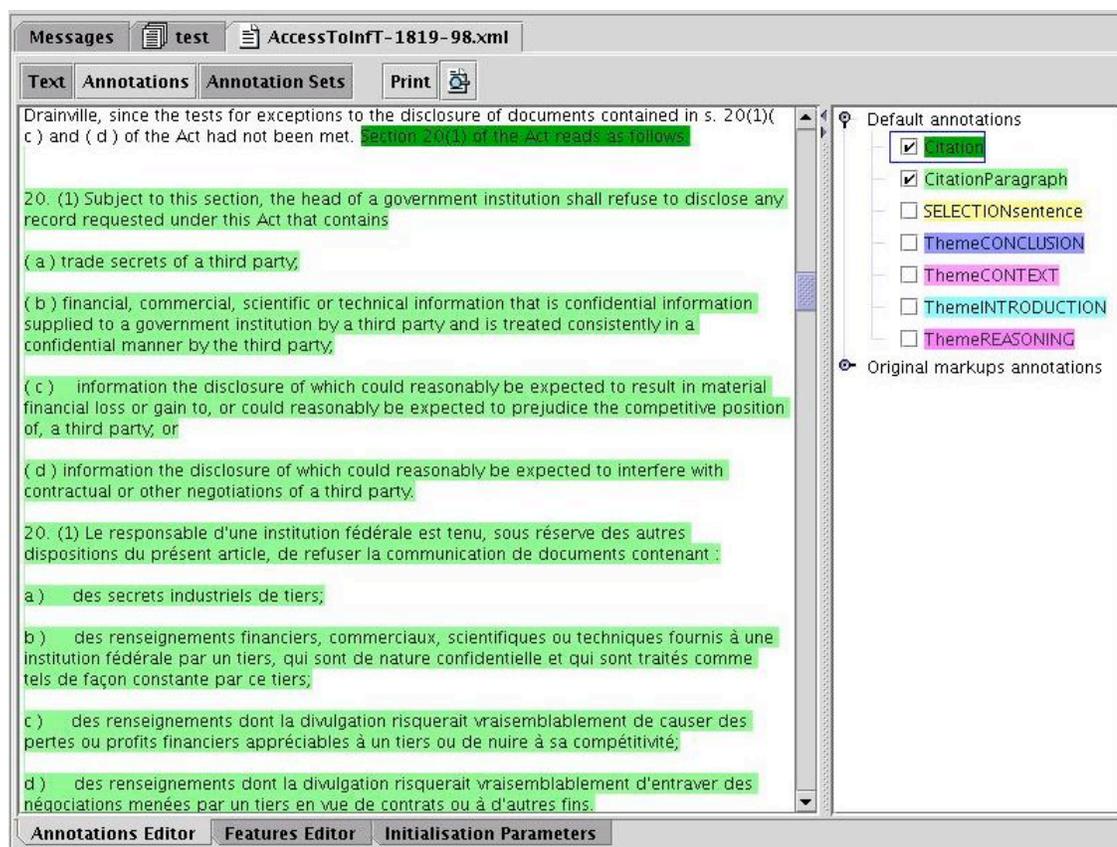


FIG. 6.2 – Sortie de module de filtrage identifiant les citations, sous l'interface de GATE

La figure 6.2 montre le résultat de l'analyse de citation par le module de filtrage, décrit à la section 6.6. L'étiquette *Citation* dans la partie droite de la figure montre les phrases annonciatrices de citation et l'étiquette *CitationParagraph* identifie les paragraphes pointés par les marqueurs de citation. Il comprend tous les éléments cités ainsi que les unités énumérées. Aucune unité identifiée comme citation n'apparaît dans le résumé.

The screenshot displays the GATE software interface. At the top, there are tabs for 'Messages', 'test', and 'AccessToInfT-1819-98.xml'. Below these are buttons for 'Text', 'Annotations', 'Annotation Sets', and 'Print'. The main text area shows several paragraphs of text, with some sentences highlighted in yellow. A table at the bottom lists the selected sentences and their features.

	Set	Start	End	Features
sentence	Default	28460	28746	{rule= SentenceSelectioncontextPhrase, kind=context}
sentence	Default	22395	22569	{rule= SentenceSelectioncontextPhrase, kind=context}
sentence	Default	18944	19027	{rule= SentenceSelectionreasoningPhrase, kind=reasoning}
sentence	Default	27614	27864	{rule= SentenceSelectionreasoningPhrase, kind=reasoning}
sentence	Default	3635	3941	{rule= SentenceSelectioncontextPhrase, kind=introduction}
sentence	Default	11413	11560	{rule= SentenceSelectionreasoningPhrase, kind=reasoning}

Annotations Editor Features Editor Initialisation Parameters

FIG. 6.3 – Sortie de module de sélection identifiant les phrases importantes, sous l'interface de GATE

La figure 6.3 montre l'extraction des phrases saillantes par le module de sélection, présenté à la section 6.7. Les unités soulignées sont des phrases pertinentes. Pour les phrases sélectionnées, on a accès aux informations sur leur position dans le document et leur valeur sémantique dans le jugement exprimée par leur thème. Par exemple pour la phrase sélectionnée, dans la fenêtre features, son thème est associé à RAISONNEMENT.

Table Style Summary	
RCMPT-979-96.html	
INTRODUCTION	[1] This is an application by Her Majesty the Queen (Crown) for an order striking out the Statement of Claim or, in the alternative, an extension of time to allow the Crown to file a Statement of Defence in the present action. [7] I believe, that before I recite the facts of the present case, it is important to note that on a motion to strike a Statement of Claim due to the fact that the Statement of Claim discloses no reasonable cause of action, it must be plain and obvious that the claim will not succeed notwithstanding the fact that the allegations in the Statement of Claim must be deemed to be true.
CONTEXT	[11] The plaintiff (Riabko) was a member of the Royal Canadian Mounted Police (RCMP) from November 6, 1978 to September 14, 1994, almost 16 years. On May 6, 1994 an Adjudication Board created under sections 43 and 44 of the According to the Crown "These actions arose from certain incidents in which the plaintiff was involved in and occurred in 1992". [13] As a result of the Board's decision of May 6, 1994, Riabko was sanctioned by requesting or ordering his resignation from the RCMP Force within 14 days. [16] On April 30, 1996, Riabko filed a Statement of Claim in this action in the Federal Court of Canada.
ISSUE	Issue[27] Does the Statement of Claim show a triable issue?
REASONING	I take this to mean that if the sections of the Act and Regulations are followed, a member may be dismissed or discharged and that the member would not be able to pursue the issue in the Courts by means of filing a Statement of Claim only alleging wrongful dismissal. [35] Because of the alleged breach of the RCMP Code of Conduct, a formal disciplinary hearing took place pursuant to section 43 of the RCMP Act, that is, an Adjudication Board was appointed to conduct a hearing into the alleged complaint. [42] It is obvious that the plaintiff Riabko did not follow the procedure set out in the RCMP Act and he is now alleging that he is claiming against Her Majesty because the process wherein he was asked to resign was an abuse of power by the Board, that is, from the very start, the process of the Board was flawed and he would thus have the right to proceed in Court. [45] I am satisfied that by having resigned, she could not avail herself of the internal process as stated in the RCMP Act and could sue for damages for sexual harassment. It must be noted that before she commenced her action before the Federal Court she did not avail herself or never took part in the process set out in the "She never did anything wrong" while in the case at bar the plaintiff was found to have contravened the RCMP Code of Conduct. [47] I am satisfied that where it cannot be shown that the power with regard to the grievance process as set out in the RCMP Act has been exceeded or abused, then there would be no cause of action. [49] I am satisfied there would be no purpose for Parliament to set out a grievance procedure by statute if a party could, after taking part in the procedure, decide to circumvent the statutory procedure.
CONCLUSION	[50] As well, after a plain reading of the Statement of Claim, and particularly paragraphs 5 and 6, I am satisfied that there is no allegation that the Adjudication Board of the RCMP abused or exceeded its jurisdiction. [51] Plaintiff's claim is struck with costs.

FIG. 6.4 – Fiche de résumé produit par LetSum, composé de 350 mots alors que le jugement source avait 4500 mots

La figure 6.4 présente une fiche de résumé produite par LetSum, comme décrit à la section 6.8. Cette fiche est en format HTML et elle est le résultat des modules de création de résumé automatique, indépendamment de l'interface de GATE. Dans cette représentation, les phrases extraites du jugement sont présentées accompagnées des rôles thématiques qu'elles jouaient dans le jugement. Ce qui permet à l'utilisateur de comprendre la nature de chaque phrase du résumé et les sens des phrases reliées à leurs thèmes. Le texte source a 4500 mots (10 pages) et le résumé a 350 mots.

6.10 Conclusion

Dans ce chapitre, nous avons présenté LetSum le système développé suivant notre méthodologie afin de résumer des jugements canadiens. Nous avons expliqué les détails de l'implémentation de chaque module de LetSum ainsi que les exemples de sorties des modules. Dans le prochain chapitre, nous présenterons l'évaluation des résumés générés par LetSum.

Chapitre 7

Évaluation des systèmes de production de résumé

L'évaluation d'un système informatique vérifie l'état d'un système afin d'en juger le niveau de qualité et aussi de l'améliorer. Dans ce chapitre, nous décrivons d'abord les concepts importants pour l'évaluation de résumés automatiques, les méthodologies d'évaluation et les stratégies des campagnes d'évaluation. Pour situer notre approche parmi les autres méthodes et mesurer la performance de LetSum, nous comparons les différentes technologies de résumé automatique et le rapport entre les résultats obtenus et les objectifs.

7.1 Introduction

Il faut qu'un résumé d'un texte en contienne les informations essentielles. Plusieurs campagnes d'évaluation de systèmes de résumé comme SUMMAC [Mani *et al.*, 1998], MUC (financé par DARPA), DUC (organisé par NIST) et le travail du workshop du *Center for Language and Speech Processing* de la Johns Hopkins University [Radev *et al.*, 2002b] ont montré l'importance de définir des mesures pour l'évaluation d'un résumé.

L'évaluation de la qualité de résumés est un problème difficile, parce qu'il n'y a pas de résumé idéal pour un texte. La comparaison avec un résumé modèle comme référence pour des résumés automatiques est très naturelle, mais des résumés rédigés par des personnes différentes ne sont pas toujours convergents au niveau du contenu. La rédaction d'un résumé demande une analyse du texte pour en dégager les idées,

les arguments, le style et les thèmes. Les rédacteurs humains dégagent les affirmations essentielles du document et les expriment dans leur propre style, ce qui donne lieu à plusieurs résumés pour le même document. Il est donc difficile de définir une métrique claire pour juger différents aspects d'un résumé comme la complétude, la thématique et la cohérence.

Spark Jones et Galliers [Spark-Jones et Galliers, 1995] ont proposé de diviser les évaluations en deux types : **intrinsèque** et **extrinsèque**. L'évaluation intrinsèque mesure les propriétés concernant la nature du sujet à évaluer et son objectif, alors que l'évaluation extrinsèque mesure les aspects concernant les impacts et les effets de sa fonction.

7.1.1 Évaluation intrinsèque

Dans le contexte du résumé, l'évaluation intrinsèque vérifie le système et la qualité du résumé lui-même. Il s'agit d'examiner la pertinence des phrases à l'intérieur même du résumé en regardant le résumé comme un objet. Pour ce genre de test, on peut comparer le contenu du résumé automatique avec un résumé modèle écrit par un humain.

L'évaluation intrinsèque peut être mesurée avec les critères suivants :

- qualité de la langue,
- lisibilité,
- cohérence entre les segments du résumé.

L'évaluation de la sortie du système peut être effectuée en comparant le résumé avec le texte original en termes de phrases pertinentes soit par jugement humain ou par des mesures statistiques. On peut mesurer la pertinence des phrases d'un résumé : avec des scores ou en les comparant avec un résumé modèle.

Pour comparer un résumé machine avec un résumé modèle, on peut calculer la précision et le rappel. La précision mesure la proportion des unités pertinentes parmi toutes les unités produites par le système. Le rappel mesure la proportion des unités pertinentes parmi tous les unités pertinentes. Ces mesures inspirées du domaine de la recherche d'information peuvent être transposées dans le domaine du résumé de la façon suivante :

P — nombre de phrases non pertinentes dans le résumé ;

Q — nombre de phrases pertinentes dans le résumé ;

R — nombre de phrases pertinentes présentes dans le résumé modèle et non trouvées

dans le résumé automatique.

Le rappel et la précision sont alors calculés comme suit :

$$\text{Rappel} = Q/(Q + R)$$

$$\text{Précision} = Q/(Q + P)$$

Les valeurs de précision et rappel sont définies entre 0 pour le pire et 1 pour le meilleur des cas. Dans le contexte du résumé automatique, un système avec une précision plus élevée signifie que la plupart des phrases extraites dans le résumé produit sont pertinentes, alors qu'un rappel élevé indique l'identification d'un nombre plus élevé de phrases pertinentes dans le texte. Un score élevé de précision et de rappel tout seuls ne signifie pas un meilleur système, parce qu'un système qui ne fournit aucune phrase aura une précision de 1 et un système qui retourne toutes les phrases du document a un rappel de 1. Pour l'évaluation, il faut considérer les deux mesures ensemble. La F-Mesure suivante est souvent utilisée comme combinaison de précision et rappel :

$$\text{F-Mesure} = (\beta^2 + 1) (\text{Précision} + \text{Rappel}) / (\beta^2 \text{Précision} + \text{Rappel})$$

Où β est le facteur de poids qui favorise la précision quand $\beta > 1$ et le rappel quand $\beta < 1$.

La métrique la plus significative pour le résumé est le rappel. Car pour une certaine taille on souhaite obtenir dans le résumé machine le nombre maximum de phrases du résumé de référence. Le principal défaut de ces critères est de postuler l'existence d'un résumé type, construit avec des phrases extraites du texte source, pouvant être utilisé comme référence.

7.1.2 Évaluation extrinsèque

L'évaluation extrinsèque examine le résumé en relation avec son utilité dans une tâche. Par exemple, pour un document qui contient les réponses à des questions prédéfinies, on demande à des lecteurs de répondre aux questions en utilisant plutôt le résumé. Si les lecteurs répondent correctement aux questions demandées, le résumé est considéré bon car il couvre les informations essentielles du document.

Lors d'une évaluation extrinsèque, on peut imaginer une liste de tâches et en mesurer le niveau de satisfaction. Par exemple, on peut mesurer l'utilité d'un résumé répondant à une requête, ou dans une tâche de question-réponse.

Pour évaluer LetSum, nous avons demandé à des juristes de mesurer la pertinence du contenu du résumé et de répondre à des questions sur la qualité du texte.

7.2 Campagnes d'évaluations

7.2.1 Évaluation de SUMMAC

Une des premières évaluations de systèmes de résumé automatique est l'évaluation TIPSTER Text Summarization Evaluation Conference (SUMMAC) [Mani *et al.*, 1998] [Mani *et al.*, 2002]. Cette campagne d'évaluation a été organisée en 1998 par le gouvernement américain avec 16 équipes participantes.

Le but de l'évaluation de SUMMAC était de juger des systèmes de résumé en termes de leur utilité dans des tâches spécifiques en utilisant 20 requêtes de TREC. Cette évaluation portait sur la pertinence d'un résumé pour 1000 documents pour les trois tâches suivantes :

- recherche d'information (ad hoc)
- catégorisation
- question-réponse

Cette évaluation comprenait des évaluations intrinsèques et extrinsèques.

Pour l'évaluation intrinsèque, les résumés automatiques ont été comparés avec des baselines et des documents originaux sur des critères de temps et F-mesure. Cette évaluation tient compte de la concordance inter-juges à l'aide du coefficient Kappa [Jean Carletta et Anderson, 1997].

Pour l'évaluation extrinsèque de la tâche ad hoc, les résumés indicatifs portaient sur les topiques essentiels des textes (utilisés en recherche d'information) à partir du critère suivant : à quel point un humain peut-il déterminer qu'un document est pertinent pour une requête en lisant son résumé ?

Dans la tâche de catégorisation, l'évaluation portait sur l'affectation d'un document à une catégorie en fonction du résumé. Pour la tâche de question-réponse, l'évaluation mesurait les réponses fournies par les résumés pour des questions portant sur les topiques des documents.

7.2.2 Évaluation de DUC

La série d'évaluation *Document Understanding Conferences*, financée par DARPA (*Defense Advanced Research Projects Agency*) réunit chaque année les chercheurs du domaine de résumé automatique afin de constater l'état des recherches et évaluer des systèmes. Au cours des évaluations intrinsèques menées pour DUC 2001, DUC 2002 et DUC 2003 les juges ont comparé manuellement les résumés produits par les machines et les résumés écrits par les humains (voir la section 1.3) en utilisant l'interface SEE (*Summary Evaluation Environment*) [Lin, 2001] qui permet de mesurer la précision et le rappel sur les scores donnés par les juges.

Lors du workshop DUC 2004 un outil d'évaluation automatique, appelé ROUGE est utilisé afin d'évaluer les résumés produits pour la tâche de production d'un résumé très court (environ 10 mots). Ce qui évalue les résumés en se basant sur la co-occurrence statistique de n-grammes.

Évaluation automatique avec ROUGE

L'évaluation manuelle des systèmes de traitement informatique du langage naturel est une tâche très lourde, car elle demande de l'expertise et du temps de la part de plusieurs évaluateurs. Afin de trouver une alternative rapide et moins chère pour l'évaluation, des méthodes statistiques ont été proposées. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [Lin et Hovy, 2003] est une méthode d'évaluation automatique de résumé qui a été utilisée pour la première fois dans la compétition de DUC 2004 comme seule mesure de fiabilité pour certaines tâches. ROUGE, basé sur le calcul statistique de co-occurrence de n-grammes est adaptée de BLEU [Papineni et Zhu, 2001], utilisé pour la traduction automatique. Cette méthode dont les résultats sont bien corrélés avec les jugements humains permet d'optimiser les systèmes et d'accélérer leur évaluation.

ROUGE comporte deux méthodes d'évaluation. ROUGE-N, dont le score est basé sur le nombre de n-grammes (normalement $1 \leq n \leq 4$) communs entre le résumé automatique et le résumé modèle. Par exemple, ROUGE-2 calcule le nombre de paires de mots successifs communs entre les résumés candidat et modèle.

La deuxième est ROUGE-L, qui considère les phrases comme une suite de séquences des mots. Cette évaluation calcule la plus longue sous-séquence commune de mots afin d'estimer la similarité entre deux résumés. La principale différence entre ROUGE et BLEU est que BLEU est basé sur la Précision alors que ROUGE est basé sur le

Rappel. Pour calculer la Précision dans BLEU, on mesure le chevauchement entre un candidat de traduction et référence, en termes de pourcentage de n-grammes communs. ROUGE compte plutôt le nombre de n-grammes communs entre le résumé machine et la référence divisé par le nombre de n-grammes dans la référence.

BLEU considère un facteur de pénalité pour une traduction plus courte que la traduction modèle ; dans ROUGE, ce facteur de pénalité est ignoré parce qu'on tronque un résumé plus long qu'une certaine limite.

Pour l'évaluation statistique d'un résumé machine, il est important de comparer le résumé avec plusieurs résumés modèles [Pastra et Saggion, 2003]. Pour cette raison, lors de l'évaluation de DUC 2004, quatre versions d'un résumé écrit par un humain étaient utilisées comme références.

7.3 Évaluation de LetSum

Nous avons mesuré la qualité des résumés de LetSum de façon intrinsèque et extrinsèque. Pour évaluer notre méthode, les résumés du système sont jugés en deux étapes : d'abord nous avons évalué les modules du système séparément (section 7.4), ensuite nous avons mesuré la qualité globale des résumés produits (section 7.5). Nous avons également comparé les résumés de LetSum avec des résumés produits par quatre autres systèmes et des résumés manuels. Pour l'évaluation des modules de LetSum, nous avons utilisé une évaluation intrinsèque à trois niveaux : la qualité des divisions en thèmes par le segmenteur thématique, la détection correcte des citations pour le module de filtrage, et le contenu des unités sélectionnées par le module de sélection et production. La comparaison entre le résumé généré par LetSum et le résumé modèle a été faite en utilisant ROUGE.

Pour l'évaluation extrinsèque de LetSum, nous avons demandé à des utilisateurs juristes de juger le contenu des résumés et leur acceptabilité. Pour chaque résumé, le recouvrement du contenu et des idées clés du document a été évalué par deux avocats.

7.4 Évaluation des modules de LetSum

Nous avons évalué les quatre modules de LetSum (voir section 5.2) séparément. Les deux premiers modules, segmentation thématique (section 5.2.1) et filtrage (section 5.2.2) sont évalués séparément, alors que les deux autres modules de sélection des

unités pertinentes (section 5.2.3) et production (section 5.2.4) ont été évalués dans le cadre de l'évaluation des résumés finals de LetSum. Nous avons comparé les sorties de module de segmentation thématique avec le corpus que nous avons annoté manuellement (avec validation d'un avocat du CRDP).

Pour l'évaluation du module de **segmentation thématique**, nous avons utilisé un corpus de test contenant 10 jugements de la cour fédérale. Ces jugements n'ont pas été utilisés pour entraîner le système, ni servi à la construction du dictionnaire des marqueurs. Pour l'évaluation de ce module les points considérés importants sont :

- la détection des thèmes,
- le degré de pertinence d'un thème pour un segment,
- le couverture des segments thématiques,
- précision des frontières entre deux thèmes.

Nous avons obtenu une précision de 100% et un rappel de 95% soit une F-mesure de 99% en choisissant $\beta = 0,5$ (importance égale aux précision et rappel). Sur 40 thèmes annotés dans le corpus, 38 thèmes ont été identifiés correctement. Dans 60% des cas, les thèmes sont retrouvés à l'aide d'un titre significatif parmi nos classes de titres (expliqués en section 5.2.1) et à l'aide des positions des segments thématiques dans le jugement. 30% des thèmes sont identifiés en utilisant des marqueurs linguistiques. Pour les 10% restants, les thèmes INTRODUCTION et CONCLUSION sont bien déterminés mais le segmenteur n'arrive pas à détecter les thèmes CONTEXTE et RAISONNEMENT. Pour régler ce problème, nous proposons une étude sur les informations temporelles que nous proposerons comme travail futur au chapitre 8.

Pour l'évaluation du module de **filtrage** de citations, nous avons utilisé 15 jugements de la cour fédérale qui n'ont pas servi à l'apprentissage du module de filtrage. Pour cette évaluation, nous avons utilisé un outil dans GATE, appelé *Annotation-Diff*, permettant d'évaluer les annotations de citation. Nous avons comparé les unités de citations identifiées par le **filtrage** avec les citations annotées manuellement dans les jugements. Le résultat de l'évaluation du module de filtrage est de 98% pour la précision et 95% pour le rappel ce qui donne 96% pour la F-mesure. Sur 60 cas de citation, 57 unités ont été identifiées correctement. Certaines citations n'étaient pas identifiées correctement à cause de la langue de rédaction des références. Dans les jugements canadiens, les juges citent parfois les références de droit telles quelles peu importe qu'elles soient en anglais ou en français. Lorsqu'il y a des marqueurs d'énumération, le système les identifie mais en l'absence des marqueurs d'énumération, il ne peut pas les distinguer.

7.4.1 Évaluation de LetSum par ROUGE

Pour le module de **sélection** et **production**, il faut mesurer les topiques extraits des documents par le système. Il est possible d’aligner automatiquement les unités de deux textes pour comparer la similarité entre les résumés modèles et les résumés produits afin de calculer la fraction du résumé modèle exprimée dans le contenu du résumé produit par le système. Pour cette évaluation des résumés de LetSum, nous avons utilisé ROUGE en les comparant avec des résumés modèles écrits par des humains. Nous avons travaillé sur les jugements de la Cour fédérale du Canada et les résumés manuels. Les sommaires sont des résumés écrits par les arrêtistes de la Cour fédérale. Les jugements sont disponibles sous le nom de *Recueil des arrêts de la Cour fédérale du Canada* sous la rubrique *Fiches analytiques*, comme nous l’avons dit en section 4.1. Pour l’évaluation de LetSum avec ROUGE, nous avons comparé 60 résumés parmi différents sujets en comparant notre système avec d’autres systèmes de résumés. Nous avons généré des résumés automatiques avec quatre systèmes :

1. Système de recherche *MEAD*, le système développé au workshop de *Center for Language and Speech Processing* de Johns Hopkins University [Radev *et al.*, 2002c], sous la direction de Dragomir R. Radev de l’Université de Michigan [Radev *et al.*, 2003],
2. Système de recherche et commercial français *Pertinence Mining*[Lehman, 1995] [Lehman, 1997],
3. Système commercial de *Microsoft Word* (option de résumé dans MS Word),
4. Une méthode *StartEnd* que nous avons définie.

Le *StartEnd* est un système basé sur les positions des segments dans le document. Nous avons mis au point cette approche suite à nos analyses du corpus des résumés manuels. Pour définir le *StartEnd* nous avons fait trois expérimentations.

D’après nos études, le début du jugement situé à la fin des DONNÉES DE LA DÉCISION (nom de la cour, lieu de l’audience, date, les références et etc.) est une partie importante qui comprend le début du thème INTRODUCTION. Nous avons défini un baseline qui prend 15% du début du texte. Ce baseline couvre des thèmes INTRODUCTION et CONTEXTE.

Un autre baseline prend 15% de la fin du jugement avant la signature du juge. Ce baseline couvre les unités des thèmes RAISONNEMENT JURIDIQUE et CONCLUSION. Nos expériences avec ROUGE ont montré que le score du premier baseline était plus élevé que le deuxième, ce qui signifie l’importance du commencement du document par rapport à sa fin.

Cette expérience, nous a conduit à définir une approche de résumé avec un taux de compression de 15%, basé sur l'algorithme suivant :

1. 8% du début du jugement et en complétant la dernière phrase si cette dernière a été coupée.
2. 4% de la fin du jugement en ajoutant la première phrase complète.

Cette dernière approche, que nous avons nommée *StartEnd* est donc assez appropriée pour les documents de style juridique même si son implémentation est assez simple. Nous avons pu utiliser un autre baseline prenant les N premières phrases de chaque thème comme résumé. N dépendant du seuil de réduction du texte source. Le problème de ce baseline est qu'il dépend de notre segmenteur thématique et il ne sera pas un baseline simple et facile à implanter. Ce baseline serait intéressant avec un bon segmenteur thématique, où justement nous voulons évaluer le module de sélection des phrases et ce dernier extrait les phrases les plus importantes dépendamment des thèmes du segment.

Grâce à l'évaluation automatique de ROUGE, nous avons pu rapidement nous positionner par rapport aux autres systèmes. Cette évaluation reste quand même assez laborieuse car dû nous avons ajusté le format du texte à résumer en fonction de chaque système. Il a ensuite fallu transformer les résumés selon format ASCII requis par ROUGE.

Les transformations que nous avons effectuées pour chaque système en considérant les textes sources en HTML, sont les suivantes :

- La version 3.07 du système *MEAD* que nous avons utilisé accepte les textes en format ASCII mais avec des balises montrant le début et la fin du document et les balises pour séparer chaque phrase du texte. Il demande aussi de créer un dossier pour chaque texte à résumer ou pour des plusieurs textes reliés au même sujet dans le cas de résumé multi-document.
- Pour le système *Pertinence Mining*, nous avons utilisé la version d'essai gratuit sur l'internet qui est limité à produire quelques résumés. Les résumés produits sont en HTML et les phrases importantes sont distinguées par un surlignage.
- Pour utiliser l'option de résumé de *Microsoft Word*, les documents originaux doivent être en format Word et les résumés produits aussi sont en Word.
- Pour fonctionner notre système *StartEnd*, nous avons éliminé des balise HTML pour avoir des textes sans balise.
- Pour LetSum, les documents entrés sont en HTML et les résumés sont en XML. Les fiches de résumés produits sont en HTML que pour cette évaluation nous avons effacé des balises HTML.

Système	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L
LetSum	0.57500	0.31381	0.20708	0.15036	0.45185
StartEnd	0.47244	0.27569	0.19391	0.14472	0.34683
MEAD	0.45581	0.22314	0.14241	0.10064	0.32089
MsWord	0.44473	0.21295	0.13747	0.09727	0.29652
Per.Mining	0.32833	0.15127	0.09798	0.07151	0.22375

TAB. 7.1 – Résultat d’évaluation intrinsèque avec ROUGE-N et ROUGE-L

- Les résumés écrits par des humains avaient aussi le format HTML, pour les sauvegarder dans le répertoire des résumés modèles de ROUGE nous les avons transformés en fichiers ASCII.

Nous avons comparé avec ROUGE les résumés de LetSum, *StartEnd* et ceux de trois autres systèmes avec les résumés humains. Pour ROUGE, nous avons utilisé l’option *stopword* qui prend en compte une liste de mots vides dans la comparaison des résumés machines et des résumés modèles, même si cette option a été ignorée pour l’évaluation de DUC 2004. Nous avons remarqué qu’en ajoutant la liste des mots vides à la comparaison, le résultat de ROUGE s’approchait plus du résultat de l’évaluation manuelle, donc dans notre expérience son résultat était plus précis.

La table 7.1 montre les résultats de l’évaluation avec ROUGE. Un score plus élevé est meilleur et indique un système plus performant. LetSum est classé au premier rang avec les meilleures notes d’évaluation. D’après cette évaluation, le deuxième système est *StartEnd*, ce qui montre l’importance de l’étude sur les documents des domaines spécifiques. Cette comparaison est illustrée à la figure 7.1.

Le fait qu’une approche simple puisse dépasser des méthodes complexes de production de résumé met en évidence les différences dans l’organisation des documents et montre aussi l’intérêt de développer un système spécifique pour un domaine. Il est de plus en plus difficile de produire un résumé général pour tous types d’utilisateurs sans prise en compte du besoin des usagers et de la tâche demandée.

7.5 Évaluation extrinsèque de LetSum

L’objectif de cette évaluation est de mesurer l’utilité du résumé automatique par rapport au résumé écrit par un humain et de comparer la qualité des résumés automatiques générés par différents systèmes. Ce test est basé sur un jugement humain.

Évaluation automatique avec ROUGE

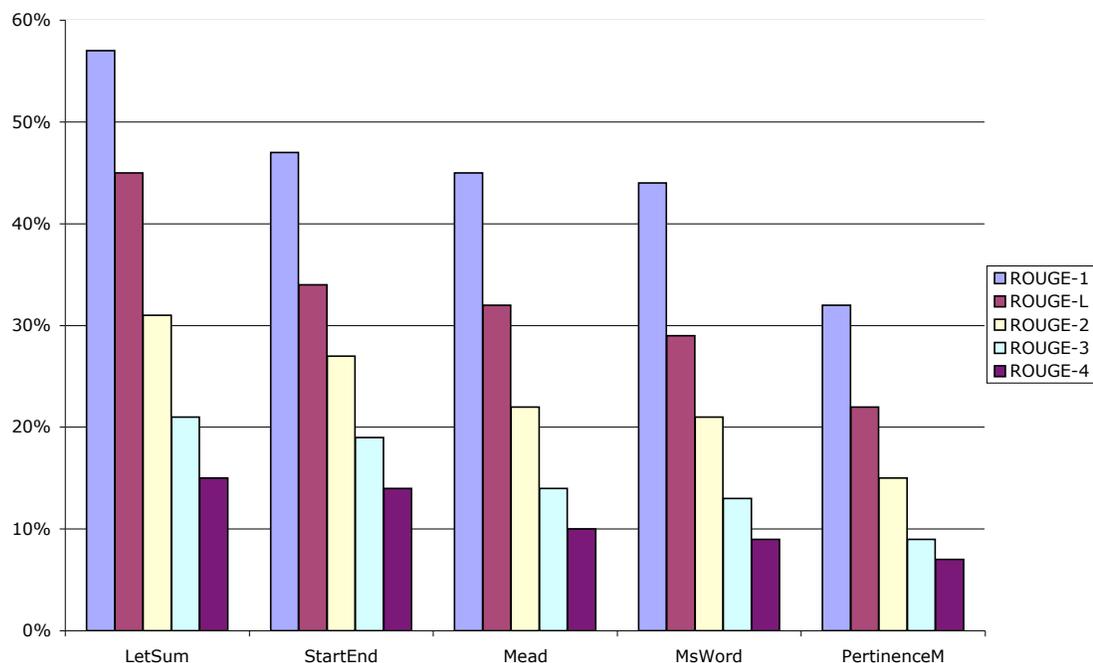


FIG. 7.1 – Graphe de comparaison des résultats d'évaluation intrinsèque avec ROUGE

7.5.1 Évaluation manuelle

L'évaluation du résumé par un humain est une démarche permettant de porter un jugement, à partir de mesures et de critères établis, sur le contenu du résumé comme la lisibilité du texte et la cohérence des segments textuels ou sur l'utilité du résumé pour des tâches spécifiques. Cette évaluation manuelle comporte une mesure à la fois qualitative et quantitative des résultats ainsi qu'un jugement de valeur concernant leur pertinence.

Ce type d'évaluation est intéressant quand le résumé modèle n'est pas disponible ou qu'une comparaison directe entre le résumé produit avec un résumé idéal n'est pas possible. En évaluant le résumé par les utilisateurs potentiels du système, on obtient une bonne mesure de la performance du système. Ce genre d'évaluation est toutefois très coûteuse, parce qu'il demande des ressources humaines et un temps considérable. Pour l'évaluation de LetSum, nous avons fait évaluer les résumés produits par LetSum

par des avocats afin d'en mesurer le bénéfice pour les juristes.

Deux dimensions sont prises en compte : contenu et qualité. Pour l'évaluation du **contenu**, nous avons défini sept points importants à retrouver dans un jugement. Si un lecteur peut déterminer les points en question en lisant le résumé, on en déduit que le résumé contient suffisamment d'informations pour couvrir les idées clés d'un jugement. Avec l'aide de M. Frédéric Pelletier, avocat au CRDP, nous avons rédigé des questions portant sur les points suivants :

- l'identification des personnes physiques ou morales qui participent au sujet d'un jugement,
- la détermination du problème en litige dans le cas,
- la précision des questions de droit soulevées dans l'affaire,
- la distinction des points sur le raisonnement du juge pour l'application des droits
- la justification de la solution
- la conclusion finale de la décision

Pour l'évaluation du contenu d'un résumé, nous avons déterminé sept questions. L'ensemble des réponses à ces questions montre le degré de couverture des idées clés du jugement source exprimées dans le résumé. La table 7.2 présente les questions juridiques utilisées lors de l'évaluation du contenu du résumé par les juristes.

Cette deuxième partie de l'expérience contient les valeurs qualitatives sur les résumés. Pour évaluer la **qualité** d'un résumé, nous avons considéré trois critères :

Lisibilité — La facilité de distinction et de perception du contenu du résumé qui en facilite la compréhension. Ce critère donne une appréciation globale du résumé. On demande si le résumé est : clair, assez clair, peu clair ou incompréhensible.

Cohérence — La présence simultanée d'éléments qui correspondent au même contenu ou qui s'accordent entre eux, qui s'harmonisent. Ce critère contient le fil conducteur du texte pour en assurer la continuité et la progression de l'information. On demande si la cohérence du texte dans le résumé est : très bonne, bonne, médiocre ou très mauvaise.

Pertinence des phrases — Caractère de ce qui est plus ou moins approprié, qui s'inscrit dans la ligne de l'objectif poursuivi. La pertinence des phrases mesure si les phrases du résumé contiennent un lien clair et direct avec le sujet dont il est question. On demande si le résumé est : très pertinent, assez pertinent, peu pertinent ou non pertinent.

L'évaluation comporte aussi une valeur d'acceptabilité sur la qualité générale du résumé. Nous avons demandé d'attribuer une valeur d'acceptabilité entre 0 et 5 pour chaque résumé (0 pour un résumé inacceptable et 5 pour un texte acceptable) sur

la qualité du texte de résumé. Les résumés avec valeur 3 jusqu'à 5 sont considérés acceptables.

7.5.2 Évaluation par les juristes

Pour cette évaluation, nous avons évalué 120 résumés en utilisant les résumés automatiques produits par cinq systèmes présentés à la section 7.4.1 et les résumés écrits par des humains. Il faut noter que dans cette évaluation, nous n'avons considéré que les textes du résumé. Nous n'avons pas généré le format tabulaire d'organisation du résumé. Nous voulions aussi normaliser l'apparence de la sortie de tous les systèmes pour ne pas influencer les juges. Ce choix pénalise toutefois LetSum car nous ne tenons pas compte de la structure thématique extraite par notre méthodologie. Les évaluateurs ne savaient pas quels résumés avaient été produits par ordinateur et lesquels avaient été écrits manuellement.

Le corpus de test contient dix jugements choisis au hasard dans différentes collections de jugements de la Cour fédérale du Canada. Nous avons généré 50 résumés automatiques et nous avons collecté 10 résumés manuels écrits par les arrêtistes de la Cour fédérale. Pour chaque résumé, nous avons répété le test deux fois, ceci nous donne deux avis par résumé. Chacun des 12 avocats du CRDP a évalué 10 résumés sur une période d'une heure.

Dans la table 7.2, nous présentons les résultats obtenus pour l'évaluation des 120 jugements où, pour chaque question, nous avons calculé le pourcentage de réponses positives données à cette question. Une réponse positive signifie que le résumé contient assez d'informations sur le point en question. Par exemple dans la deuxième colonne, les résumés produits par la moyenne de toutes les méthodes ont couvert les informations sur la présentation des parties en litige (Q1) dans 47% des cas.

Les résultats montrent que le résumé de style télégraphique rédigé par les arrêtistes, la moyenne pour Humain dans la table 7.2, contient 80% des informations clés d'un jugement. Pour toutes les questions, sauf les questions 1 et 4, le résumé humain contient plus de 80% des réponses correctes. Les réponses des avocats à la première question concernant l'identification des parties est de 55%, ce qui signifie que dans le résumé d'un jugement, on ne cherche pas à savoir qui sont les parties mais plutôt quel est le sujet du jugement. Le résumé de l'analyse du juge et la conclusion. La quatrième question relative à l'application du droit aux faits par le juge est en seconde place en importance par rapport au raisonnement du juge sur les questions de droit (question

Après avoir lu le résumé peut-on déterminer :

- Q1. Qui sont les parties en litige ?
 Q2. Quel est le problème en litige ?
 Q3. Les questions de droit soulevées ?
 Q4. Comment le juge a appliqué le droit aux faits ?
 Q5. Les motifs couvrent-ils les questions de droit ?
 Q6. Le résumé contient-il les motifs déterminants pour arriver à la conclusion ?
 Q7. Le résultat final de la cour ?

Résumé	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Moyenne
Humain	55,00	90,00	90,00	70,00	80,00	85,00	95,00	80,71
LetSum	50,00	90,00	80,00	75,00	75,00	85,00	85,00	77,14
StartEnd	65,00	100,00	100,00	70,00	80,00	70,00	90,00	82,14
MEAD	55,00	100,00	95,00	65,00	50,00	50,00	40,00	65,00
MsWord	30,00	80,00	85,00	60,00	45,00	45,00	60,00	57,86
Per.Mining	25,00	65,00	55,00	35,00	35,00	35,00	45,00	42,14
Moyenne	46,67	87,50	84,17	62,50	60,83	61,67	69,17	67,50

TAB. 7.2 – Questions juridiques utilisées lors de l'évaluation du contenu du résumé par les juristes et les résultats d'évaluation extrinsèque, les pourcentages des réponses positives pour les sept questions juridiques

5) et le motif déterminant pour arriver à la conclusion (question 6).

Les quatre questions du thème de RAISONNEMENT (Q4, Q5 et Q6) dans l'analyse individuelle d'un jugement, selon la confirmation des juristes sont très pertinentes, mais l'évaluation montre que pour une lecture rapide d'un résumé, ces questions semblent moins importantes. Pour cette raison la méthode de *StartEnd* a donné des bons résultats.

LetSum a donc très bien répondu aux exigences des avocats pour des résumés automatiques. Ses résultats sont très proches de ceux des résumés manuels et sa performance est supérieure à celle des autres systèmes commerciaux *Microsoft Word* et *Pertinence Mining*, y compris le système de recherche *MEAD*. La figure 7.2 compare les résultats de l'évaluation extrinsèque pour les questions juridiques.

Notre méthode *StartEnd*, basée sur la position des segments, a également donné de bons résultats. Notre heuristique pour les positions des segments était appropriée, même si elle diffère du baseline utilisé normalement pour les articles journaux. Par le comportement du système *MEAD*, spécialisé pour les articles des journaux, on peut voir que les questions qui possèdent les réponses placées au début du document sont

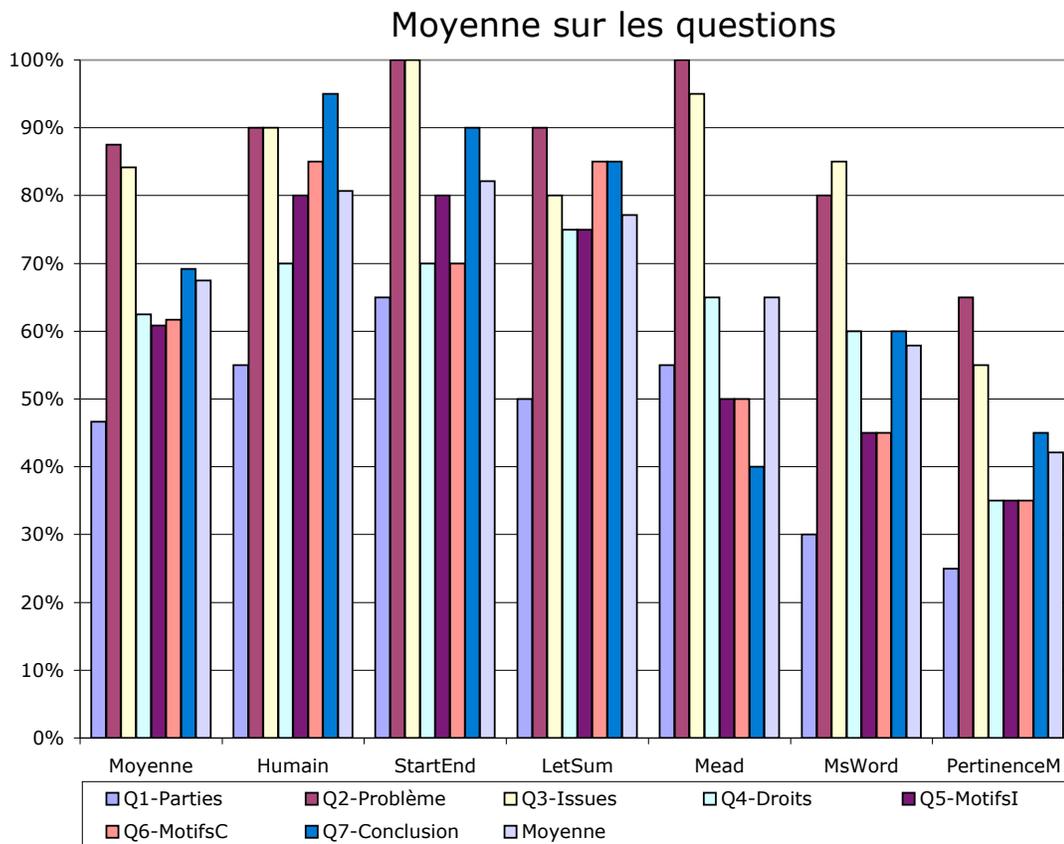


FIG. 7.2 – Graphe de comparaison des résultats d'évaluation extrinsèque sur les questions juridiques

bien répondues alors que le recouvrement des informations clés sur les questions avec réponses dans d'autres positions dans le texte n'est pas satisfaisant. Les systèmes commerciaux comme *Microsoft Word* et *Pertinence Mining* ont les scores les plus faibles dans l'évaluation, car ils produisent des résumés génériques qui ne satisfont pas vraiment les utilisateurs dans un domaine spécifique comme le droit.

La table 7.3 montre les résultats de l'évaluation de la qualité du résumé. Pour les trois critères, lisibilité, cohérence et pertinence des phrases du résumé, les valeurs sont entre 0 et 3 (section 7.5.1). La qualité des résumés produits par LetSum est supérieure à celle des autres méthodes. La lisibilité du résumé de LetSum est jugée clair, la cohérence est évaluée très bonne et les pertinences des phrases sont jugées très pertinentes pour les besoins des avocats. Les condensés rédigés par un humain sont jugés bons en cohérence (et non pas très bons) parce qu'ils sont en style télégraphique alors que LetSum et les

Résumé	Lisibilité 0-3	Cohérence 0-3	Pertinence 0-3	Acceptabilité 0-5
Humain	2,00	1,95	2,15	3,43
LetSum	2,30	2,30	2,25	3,43
StartEnd	2,40	2,20	2,10	3,68
MEAD	2,15	1,90	2,05	3,23
MsWord	1,65	1,25	1,60	2,63
Per.Mining	1,40	1,10	1,40	2,23
Moyenne	1,98	1,78	1,93	3,10

TAB. 7.3 – Résultats d’évaluation extrinsèque selon les valeurs qualitatives entre 0 et 3 sur lisibilité, cohérence et pertinence des phrases, valeur d’acceptabilité est entre 0 et 5 sur la qualité générale du résumé

autres systèmes font de l’extraction de phrases. La figure 7.3 montre les comparaisons visuelles des résultats de l’évaluation pour les valeurs qualitatives.

Du point de vue de l’acceptabilité du résumé, les résumés de LetSum sont jugés de niveau équivalent à celui des résumés écrits par les arrêtières des cours. Encore une fois la méthode de positions des phrases dans le jugement donne de très bons scores pour ce critère d’évaluation. Cette comparaison est illustrée à la figure 7.4.

Il faut noter que dans cette partie de l’évaluation il y a peu de différence entre le système StartEnd et LetSum, un système nettement plus élaboré. Ceci peut en partie s’expliquer par le fait que nous n’avons pas considéré le format tabulaire produit par LetSum basé sur l’analyse thématique du texte qui distingue notre méthode.

7.6 Résumé

Nous avons présenté deux types d’évaluation des résumés produits par LetSum que nous avons comparés avec différentes technologies de production automatique de résumé. Les moyennes des scores pour les phrases extraites par LetSum sont jugées meilleures que les phrases sélectionnées par les autres systèmes. Les résumés de LetSum ont été jugés par les juristes de qualité comparable avec celle des résumés rédigés manuellement par les spécialistes du domaine juridique. Ces résultats montrent l’efficacité de notre méthode et nous encouragent à continuer de travailler sur différents types de jugements dans le domaine juridique. Dans le prochain chapitre, nous présenterons de nouvelles perspectives et nos travaux futurs sur ce sujet de recherche.

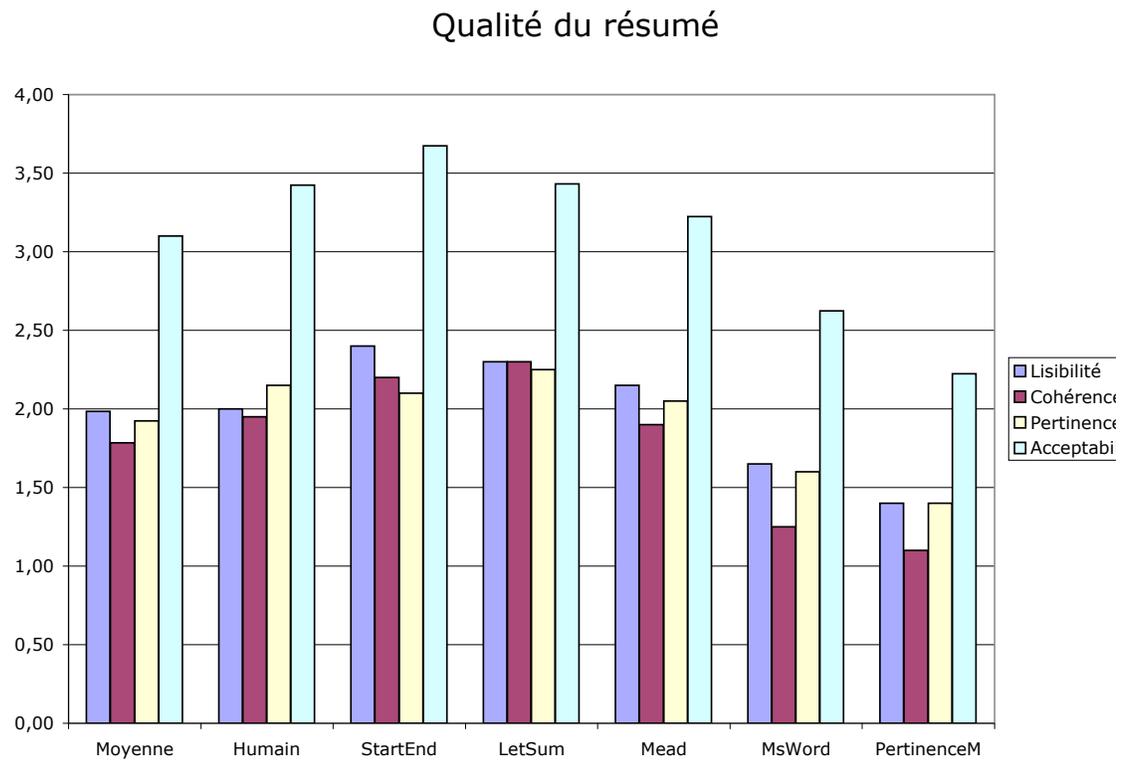


FIG. 7.3 – Graphe de comparaison des résultats d'évaluation pour les valeurs qualitatives

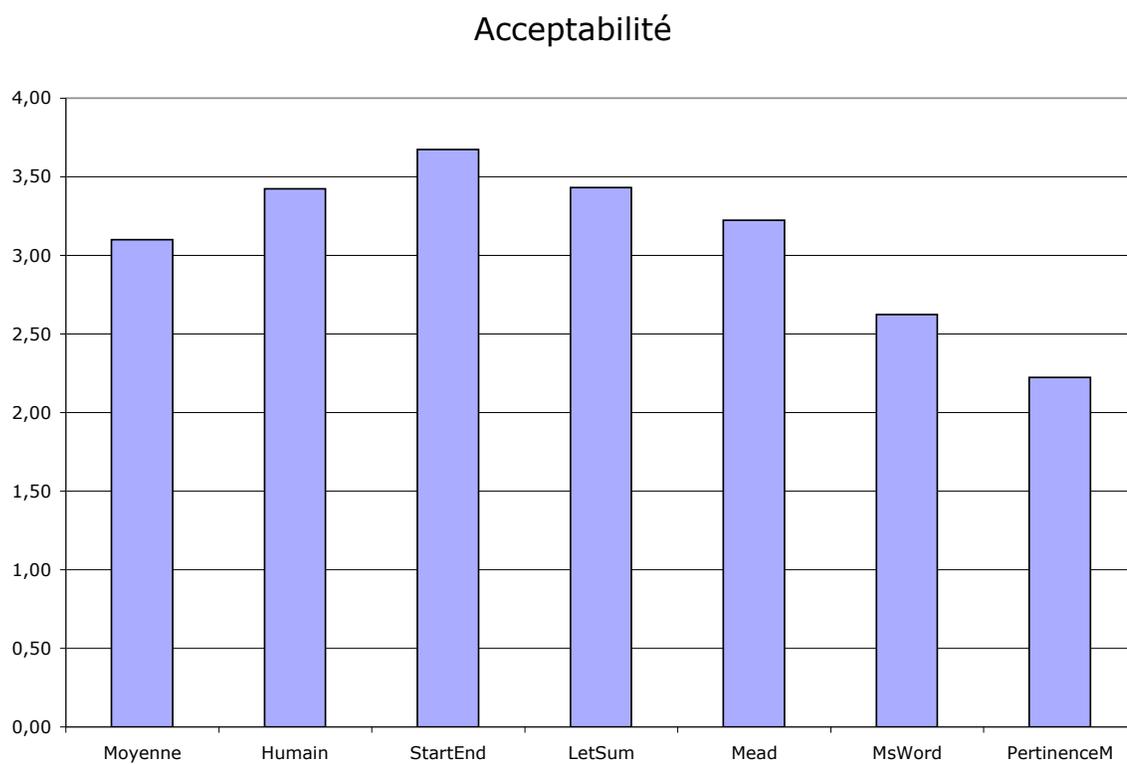


FIG. 7.4 – Graphe de comparaison des résultats d'évaluation pour la valeur d'acceptabilité sur la qualité générale du résumé

Chapitre 8

Perspectives futures

Dans ce chapitre, nous proposons d'autres travaux qui pourraient être effectués suite à nos travaux, de façon à produire des résumés très courts, à y intégrer des aspects temporels et à adapter notre méthode pour la production de résumé des autres cours de justice et aussi pour le traitement des jugements en français.

8.1 Génération de titres significatifs pour les jugements

Edmundson a suggéré [Edmundson, 1969] que les principaux thèmes du texte sont véhiculés dans les titres et que les phrases importantes sont celles qui contiennent ces mots. Donc, les scores des phrases dépendant de l'apparition de mots des titres et de leur regroupement.

Lors de DUC 2002, [Daumé *et al.*, 2002] a proposé la compression des phrases pour la production de résumé en utilisant le *canal bruité*. Une extension de ce travail est décrite par [Daumé et Marcu, 2004] qui étudie la possibilité de fusionner deux phrases. Dans DUC 2003, cette idée était une tâche additionnelle dans la compétition sous forme de production de titre (*headline generation*) pour les articles journalistiques sous le nom de résumé très court (*very short summary*). Dorr [Dorr *et al.*, 2003] a expérimenté l'analyse syntaxique de la première phrase du texte, mais les résultats de l'évaluation de DUC 2003 ont montré que le meilleur système était le système de Chali [Chali *et al.*, 2003] qui a identifié une phrase comme étant le titre de l'article avec le maximum de connexions lexicales [Barzilay et Elhadad, 1997] dans le segment thématique. Afin de diviser les textes de style journalistique le segmenteur C99 a été

utilisé [Choi, 2000].

Dans les articles journalistiques, les messages les plus importants apparaissent dans les mots de titre. Mais dans les jugements, il n’y a pas de véritable titre pour le document. Par contre les premiers paragraphes du jugement donnent un aperçu global sur le cas. À partir de premier thème, que nous appelons *Introduction* dans la fiche de résumé, nous pourrions générer un titre informatif pour le jugement limité à un maximum de 15 mots. Ce titre sera un résumé très court pour le jugement.

Une application des titres significatifs est la présentation des résultats de la recherche d’informations. Pour une requête dans la base de données de CanLII, des dizaines de documents peuvent être présentés avec leurs titres significatifs. En regardant les titres des jugements, qui montrent le but de chaque document, l’utilisateur pourra décider rapidement de la pertinence d’une jurisprudence pour sa recherche.

Pour cette tâche, nous avons expérimenté l’analyseur syntaxique *Link Grammar*¹ qui donne l’arbre de l’analyse morpho-syntaxique correspondant aux phrases du texte source. Nous avons analysé les arbres syntaxiques pour identifier les parties moins importantes de la phrase. Nous proposons l’algorithme suivant pour conserver ou supprimer les unités textuelles dans une phrase et pour ne garder qu’environ 15 mots. La table 8.1 montre un exemple de cette opération.

Dans la structure syntaxique et hiérarchique des phrases,

- conserver les premiers niveaux des groupes nominaux, des unités étiquetées NP,
- conserver les premiers niveaux des groupes verbaux, des unités textuelles étiquetées VP,
- conserver certaines subordonnées relatives, des unités textuelles étiquetées S-BAR,
- supprimer les niveaux plus profonds des suites de mots après une préposition étiquetés PP,
- supprimer certaines expressions temporelles,
- supprimer certains déterminants.

Au début de notre projet, nous avons expérimenté la production d’un titre significatif, mais nous avons réalisé que, sans une étude du plan thématique du jugement, nous n’arrivions pas à de bons résultats. Nous pourrions maintenant reprendre cette idée, parce qu’avec les connaissances sur la structure du jugement, la réalisation de cette étape ne serait pas difficile. D’autant plus que nos partenaires de CanLII ont montré un grand intérêt pour ce type de *very short summary*.

¹<http://www.link.cs.cmu.edu/link/>

<p><i>Entrée</i> : The applicant, Rogers Communications Inc., declares that Order-in-Council 1997-486 dated and registered with the clerk of the Privy Council on April 8, 1997, is invalid as being ultra vires the Broadcasting Act.</p>
<p>(S (NP The applicant), (NP Rogers Communications Inc.), (VP declares (SBAR that (NP Order-in-Council 1997-486) (VP dated and registered (PP with the clerk of the Privy Council) (PP on April 8 , 1997)),) (VP is invalid as being ultra vires the Broadcasting Act.)))</p>
<p><i>Sortie</i> : The applicant declares that Order-in-Council 1997-486, is invalid as being ultra vires the Broadcasting Act.</p>

TAB. 8.1 – Génération d’un titre informatif pour un jugement d’environ 15 mots. Nous conservons les premiers niveaux des groupes nominaux, qui sont des unités textuelles étiquetées NP, les premiers niveaux des groupes verbaux, avec étiquetés VP, les subordinés relatives étiquetées S-BAR commençant par *that*.

8.2 Aspects temporels dans le résumé

Nous avons montré l’importance de l’identification de la structure thématique des jugements, mais il serait important d’en augmenter la performance. Les marqueurs linguistiques au niveau des sens et de leurs compositions dans le document étant limités dans les d’informations qu’ils fournissent sur les thèmes, nous comptons aborder de nouvelles dimensions : le temps et l’aspect.

Pour développer LetSum, nous avons construit un dictionnaire de concepts saillants et de règles sémantiques pour la créer des résumés ; l’étape suivante sera d’élargir notre champ de recherche à la dimension des aspects temporels dans le résumé.

Les informations temporelles ont une importance considérable dans la langue naturelle. La quantité des travaux de recherche pour élaborer les approches théoriques afin de modéliser le temps et l’aspect montre la place privilégiée des informations temporelles pour l’identification des structures des événements dans le traitement automatique des langues (TAL). L’extraction des informations temporelles à analyser demande des outils d’annotations automatiques pour identifier les indicateurs reliés au temps. En nous basant sur une analyse des informations temporelles du document, nous pourrions obtenir des informations plus précises sur la sémantique du plan thématique

du jugement. Dans notre thèse, nous avons déjà fait ressortir le fait que dans le segment CONTEXTE les événements sont décrits plutôt au passé alors que dans le thème RAISONNEMENT JURIDIQUE les marqueurs sont au présent.

Pour l'analyse des segments thématiques pour en extraire les informations saillantes en fonction des besoins des juristes, il est important d'extraire les entités nommées comme le lieu du crime ou les dates des événements. Le premier problème est que, dans un document juridique, l'interprétation sémantique des phrases peut être changée selon ces expressions temporelles. Ce qui conduit à l'analyse des liens entre les expressions temporelles et les événements. Par exemple l'importance d'une date dans le résumé, une date de naissance n'est pas équivalente avec celle de l'explosion d'une bombe qui peut faire condamner une personne. Il est nécessaire de caractériser les expressions temporelles pour savoir quelles sortes d'informations peuvent être extraites avec ces expressions. En considérant ces caractéristiques, on pourrait retrouver les événements reliés aux expressions temporelles en question.

Un autre indice important est la relation temporelle. Une fois les indicateurs annotés de façon fiable, il faut retrouver un repère de temps permettant de situer les autres. Ceci demande une analyse des relations entre les expressions et les événements attachés. Le problème est de retrouver les points de référence des expressions temporelles indirectes (ex. *dans ce mois, au début du mois*) pour identifier l'ordre des événements. L'extraction des expressions temporelles et de leurs références nous aident à retrouver les événements les plus marquants dans un jugement et à présenter à l'utilisateur des événements pertinents dans l'ordre chronologique.

Dans notre thèse, nous avons utilisé la sélection des phrases du document en fonction du plan thématique du jugement. La dimension temporelle peut servir comme extension des connaissances sémantiques des textes de jurisprudence. Ces informations peuvent augmenter la performance du segmenteur thématique de LetSum qui divise le jugement en thèmes principaux. Dans LetSum, nous avons pondéré les phrases du texte selon le sens des marqueurs linguistiques et des vocabulaires juridiques, alors que les informations temporelles permettent d'analyser les structures détaillées des événements pour chaque thème dans le jugement. Par exemple, le texte narratif du thème CONTEXTE peut être analysé grâce aux caractéristiques suivantes :

- présence d'une suite d'événements qui font partie d'un fait ;
- présence des dates complètes des événements ;
- présence de liens logiques de temps (*then, following day, ...*) ;
- temps des verbes (en général au passé) ;
- présence des verbes d'action.

L'analyse temporelle du document dans LetSum pourrait raffiner le filtrage d'information en fonction du profil de l'utilisateur. Par exemple, la construction d'un résumé sur un point précis dans le jugement. Pour cette tâche, il faudra regrouper les informations reliées en ordre chronologique et donner priorité aux événements importants.

8.3 Production de résumé pour les autres types de jugement

Nous avons étudié les jugements de la Cour fédérale de première instance, mais il reste à élaborer la recherche sur d'autres types de jugement. Notre approche de la constitution des fiches de résumés devrait rester la même pour d'autres cours de justice, mais les détails des modules du système demanderont probablement une adaptation. Par exemple, dans les jugements de la Cour fédérale que nous avons étudié, le thème CONTEXTE présente les parties et leurs demandes. La détermination des faits y est importante. Mais dans la *Cour d'Appel*, le thème CONTEXTE contient plutôt le histoire judiciaire du procès, car un jugement de la Cour d'Appel est déjà passé par plusieurs cours de niveau inférieur. Dans les jugements de cette catégorie, le juge cite les raisonnements des autres juges sur le procès et leurs interprétations et, dans son analyse, il prononce son avis par rapport aux jugements précédents. Donc il faudra ajouter les indicateurs adaptés à ce type de phrases. Ce qu'il faut aussi ajouter à LetSum sera dans le module de filtrage pour éliminer les segments moins importants de citation. Dans les jugements de la Cour d'Appel, les citations portent sur les articles des lois plus des citations des raisonnements des autres juges qui ont analysé le problème en litige lors des cours de niveau inférieur.

8.4 Traitement des jugements en français

Notre méthodologie de création de résumés automatiques a été développée pour les jugements en anglais, or nous souhaitons l'adapter pour les documents en français. Les principaux composants du système devraient rester les mêmes, mais il faudrait adapter le dictionnaire des marqueurs linguistiques. Nous avons l'avantage d'avoir un corpus bilingue des jugements en anglais et français, ce qui nous donne la possibilité d'aligner les textes en parallèle et d'extraire les termes français correspondant aux termes anglais que nous utilisons actuellement. Nous développerons aussi les règles sémantiques en français pour vérifier la présence de marqueurs dans le jugement.

Chapitre 9

Conclusion

Dans cette thèse, nous avons étudié le problème de la production automatique de résumés pour les textes juridiques. Notre approche au résumé automatique a l'avantage de fournir des moyens de concevoir des résumés courts pour différents types d'utilisateurs des documents juridiques : des étudiants jusqu'aux avocats et aux juges.

Le domaine juridique est un vaste domaine avec un grand besoin pour le résumé automatique. Au Canada, il y a 70 000 avocats dont un peu moins de la moitié font de la recherche juridique. Soit environ 30 000 avocats au Canada et plus de 300 000 avocats aux USA sont susceptibles de rechercher de la jurisprudence. Toutes les synthèses des jurisprudences font manuellement par les juristes. Nous avons discuté avec des avocats pratiquant dans des cabinets d'avocats au Québec et en Ontario sur leurs méthodes d'analyse de jugements. Après un processus d'analyse d'un dossier un avocat-rechercheur cherche d'abord les jurisprudences potentiellement pertinentes. Le juriste peut alors explorer CanLII pour trouver d'autres cas similaires au cas étudié.

Lors qu'un résumé est disponible, le juriste a une idée du contenu de la décision il est plus facile de savoir si elle a un potentiel de pertinence. Chaque résumé peut sauver, dans la consultation d'une liste de résultats de recherche, deux minutes à la personne qui fait la recherche. Une recherche typique dans CanLII donne plus de trente résultats, on pourrait donc sauver une heure environ. Comme un avocat-rechercheur facture au moins 100\$ de l'heure à son client, et que plusieurs recherches peuvent être requises par dossier, par exemple pour 20 recherches il y aura 20 heures d'économies, donc 2 000\$ sur un seul cas. Dans la deuxième étape d'analyse d'un dossier juridique, une fois certaines jurisprudences sélectionnées, les avocats étudient manuellement le texte pour le diviser le jugement en différentes parties : l'introduction sur le contexte du litige, l'avis des avocats des parties, le raisonnement du juge, les droits applicables

sur le cas et les références cités, trouver le ratio ou le motif du jugement qui est le coeur de l'analyse du juge et qu'il conduit le litige à une solution et finalement le résultat de la décision du juge.

L'analyse d'un jugement, environ 10 pages, prend des heures de travail par ce qu'il nécessite de lire le texte, de le comprendre, de l'interprétation et de le séparer en thèmes, de faire les liens entre les éléments mentionnés dans le jugement, de supprimer les parties non pertinentes et de sélectionner les phrases importantes. Pour chaque question ou problème présent dans le dossier, les avocats font la copier-coller des phrases du jugement avec la référence du jugement et ils les présentent aux tribunaux. Ceci est la méthode traditionnelle des analyses des dossiers en Amérique du nord pratiquée par des milliers des juristes. Notre approche de résumé simule dans les faits cette méthode d'analyse manuelle par les experts, car elle se base sur l'extraction des phrases dépendant de la structure thématique du jugement.

Dans notre recherche, nous avons considéré le domaine juridique comme un domaine restreint mais un grand champ à explorer pour le traitement automatique des langues. Le système judiciaire est un des plus importants organismes au niveau gouvernemental, mais les moyens pour l'informatiser n'ont commencé que récemment. Nous avons constaté que, malgré les limites des méthodes de TAL, nous sommes capable de créer un outil puissant pour aider les juristes. L'utilisation des résumés automatiques économise du temps, des coûts et des expertises. Ces économies de ressources protègent les intérêts du gouvernement et des peuples en tant que des clients attendant de recevoir un service juridique.

Notre recherche est effectuée dans un contexte pratique en travaillant directement avec les experts du domaine juridique. Nous avons développé une méthode pour automatiser la production de résumé à travers une analyse fine des différents aspects d'un jugement. Nous avons élaboré une approche de résumé fondée sur l'idée que l'importance d'une phrase dépend de son thème dans le document. En nous basant sur le plan organisationnel de jurisprudence proposé par la juge Mailhot [Mailhot, 1998] et nos analyses de corpus, nous avons catégorisé le jugement en sept thèmes principaux : DONNÉES DE LA DÉCISION, INTRODUCTION, CONTEXTE, SOUMISSION, ISSUES, RAISONNEMENT JURIDIQUE et CONCLUSION. Nous avons également identifié les segments de citation.

Les phrases extraites apparaissent sous la forme d'une fiche de résumé. Notre méthode vise la cohérence dans les extraits produits automatiquement tout en augmentant la lisibilité. Elle se base sur des consignes fournies dans une fiche de résumé. Cette méthode consiste à extraire du texte source une liste ordonnée et à remplir

les champs reliés aux thèmes du jugement avec ces extraits. LetSum, notre prototype développé suivant cette méthodologie, a été évalué par des avocats qui sont les utilisateurs potentiels du système. Les résultats de l'évaluation des résumés de LetSum a montré une qualité proche des celles des résumés écrits par des humains.

Cette étude a été effectuée dans un contexte de production de résumé sur une grande quantité de textes juridiques comme la bibliothèque numérique de droit de CanLII. Durant notre thèse, nous avons collaboré avec les avocats pour avoir leurs suggestions et pour valider les différents aspects de notre recherche.

Dans les différentes étapes de notre étude, nous avons cherché à maximiser la précision de notre analyse en vue de diminuer les erreurs, car les textes de lois sont très précieux. L'excellente évaluation de LetSum est le témoin de la validité de notre approche. En faisant ressortir les points essentiels des jugements, nous espérons avoir rendu la justice plus accessible à tous et aussi aider la société.

Il est important de mentionner qu'on est encore loin de produire des résumés comparables aux résumés produits par un humain mais en considérant les efforts et les ressources nécessaires pour écrire les résumés manuels notre système peut être très utile comme un outil pratique pour les juristes.

Bibliographie

- [Barzilay et Elhadad, 1997] Regina Barzilay et M. Elhadad. Using Lexical Chains for Text Summarization. In *Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization*, pages 10–17, Madrid, Spain, July 1997.
- [Baxendale, 1958] P. B. Baxendale. Man-Made Index for Technical Literature- an Experiment. *IBM Journal of Research and Development*, 2(4) :354–361, 1958.
- [Borko et Bernier, 1975] Harold Borko et Charles Bernier. *Abstracting Concepts and Methods*. Academic Press, New York, 1975.
- [Chali *et al.*, 2003] Yllias Chali, Kolla Maheedhar, Singh Nanak et Zhenshuan Zhang. The University of Lethbridge Text Summarizer at DUC 2003. In *Document Understanding Conference (DUC'2003)*, pages 148–152, Edmonton, Alberta, Canada, 2003.
- [Choi, 2000] Freddy Choi. Advances in domain independent linear text segmentation. In *Proceeding of the 1 st North American Chapter of the Association for Computational Linguistics*, pages 26–33, Seattle, Washington, 2000.
- [Cremmins, 1996] Edward T. Cremmins. *The Art of Abstracting*. Information Resources Press, Arlington, VA, 2nd édition, 1996.
- [Cunningham *et al.*, 2002] H. Cunningham, D. Maynard, K. Bontcheva et V. Tablan. GATE : A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, Philadelphia, July 2002.
- [Daumé *et al.*, 2002] Hal Daumé, Daniel Marcu, Abdessamad Echihabi, Dragos Stefan Munteanu et Radu Soricut. GLEANS : A Generator of Logical Extracts and Abstracts for Nice Summaries. In *Document Understanding Conference (DUC'2002)*, pages 9–14, Philadelphia, PA, July 2002.
- [Daumé et Marcu, 2004] Hal Daumé et Daniel Marcu. Sentence Fusion is an Ill-Defined Summarization Task. In *the Text Summarization Branches Out Workshop at ACL*, pages 96–103, Barcelona, Spain, July 25 – 26 2004.
- [Desclés *et al.*, 1991] J.-P. Desclés, C. Jouis, H.-G. Oh et D. Maire Reppert. *Chapitre Exploration contextuelle et sémantique : un système expert qui trouve les valeurs sémantiques des temps de l'indicatif dans un texte*. D. Hérin-Aime, R. Dieng, J-P. Regourd, J-P. Angoujard (eds.), IOS Press, Amsterdam, 1991.
- [Desclés, 1988] Jean-Pierre Desclés. Langage et cognition. *RSSI, Association Canadienne de Sémiotique*, 8(2), 1988.

- [Desclés, 1997] Jean-Pierre Desclés. Systèmes de l'exploration contextuelle. *Co-texte et calcul du sens, (Claude Guimier)*. Presses de l'universitaires de Caen, pages 215–232, 1997.
- [Dorr *et al.*, 2003] Bonnie Dorr, David Zajic et Richard Schwartz. Hedge Trimmer : A Parse-and-Trim Approach to Headline Generation. In *HLT-NAACL 2003 Workshop : Text Summarization (DUC03)*, Dragomir Radev et Simone Teufel, éditeurs, pages 1–8, Edmonton, Alberta, Canada, May 31 - June 1 2003. Association for Computational Linguistics.
- [Douzidia Soufiane et Lapalme, 2004] Fouad Douzidia Soufiane et Guy Lapalme. Lakhass, an Arabic summarization system. In *DUC04 : NAACL'2004 Workshop in Automatic Text Summarization*, Boston, MA, USA, May 2-7 2004.
- [DUC, 2004] DUC. Document Understanding Conference, May 2004.
- [Edmundson, 1969] H. P. Edmundson. New Methods in Automatic Extracting. *Journal of the Association for Computing Machinery*, 16(2) :264–285, April 1969.
- [Farzindar *et al.*, 2002a] Atefeh Farzindar, Guy Lapalme et Horacio Saggion. Evaluation à DUC2002 d'une adaptation de SumUM, un système de résumé automatique. In *Association pour le Traitement Automatique des Langues (ATALA) Le résumé de texte automatique : solutions et perspectives*, Paris, France, décembre 14 2002.
- [Farzindar *et al.*, 2002b] Atefeh Farzindar, Guy Lapalme et Horacio Saggion. Summaries with SumUM and its Expansion for Document Understanding Conference. In *DUC02 : ACL'2002 Workshop in Automatic Text Summarization*, Philadelphia, U.S.A., July 6–12 2002.
- [Farzindar *et al.*, 2004] Atefeh Farzindar, Guy Lapalme et Jean-Pierre Desclés. Résumé de textes juridiques par identification de leur structure thématique. *Traitement Automatique des Langues (TAL), Numéro spécial sur : Le résumé automatique de texte : solutions et perspectives*, 45(1) :39–65, 2004.
- [Farzindar et Lapalme, 2003] Atefeh Farzindar et Guy Lapalme. Using Background Information for Multi-Document Summarization and Summaries in Response to a Question. In *DUC03 : NAACL'2003 Workshop in Automatic Text Summarization*, pages 168–173, Edmonton, Alberta, Canada, May 31 - June 1 2003.
- [Farzindar et Lapalme, 2004a] Atefeh Farzindar et Guy Lapalme. Legal text summarization by exploration of the thematic structures and argumentative roles. In *Text Summarization Branches Out Workshop held in conjunction with ACL'2004*, pages 27–34, Barcelona, Spain, 25–26 July 2004.
- [Farzindar et Lapalme, 2004b] Atefeh Farzindar et Guy Lapalme. LetSum, a Text Summarization system in Law field. In *THE FACE OF TEXT conference (Computer Assisted Text Analysis in the Humanities)*, pages 27–36, McMaster University, Hamilton, Ontario, Canada, 19-21 November 2004.
- [Farzindar et Lapalme, 2004c] Atefeh Farzindar et Guy Lapalme. LetSum, an automatic Legal Text Summarizing system. In *Legal Knowledge and Information Systems, JURIX 2004 : the Seventeenth Annual Conference*, Thomas F. Gordon, éditeur, pages 11–18. IOS Press, Berlin, Germany, 8-10 December 2004.
- [Farzindar et Lapalme, 2004d] Atefeh Farzindar et Guy Lapalme. The Use of Thematic Structure and Concept Identification for Legal Text Summarization. In *Computatio-*

- nal Linguistics in the North-East (CLiNE2002)*, pages 67–71, Concordia University, Montréal, Québec, Canada, August 2004.
- [Farzindar et Lapalme, 2005] Atefeh Farzindar et Guy Lapalme. Production automatique du résumé de textes juridiques : évaluation de qualité et d’acceptabilité. In *Traitement Automatique des Langues Naturelles (TALN 2005)*, Dourdan, France, 6-10 Juin 2005.
- [Farzindar, 2004] Atefeh Farzindar. Développement d’un système de résumé automatique de textes juridiques. In *TALN-RECITAL’2004*, pages 39–44, Fès, Maroc, 19-22 Avril 2004.
- [Gelbart et Smith, 1991] Daphne Gelbart et J. C. Smith. Beyond Boolean search, Flexicon, a legal text-based intelligent system. In *the Third International Conference on Artificial Intelligence and Law*, New York, U.S.A., 1991.
- [Grover *et al.*, 2003] Claire Grover, Ben Hachey et Chris Korycinski. Summarising Legal Texts : Sentential Tense and Argumentative Roles. In *HLT-NAACL 2003 Workshop : Text Summarization (DUC03)*, Dragomir Radev et Simone Teufel, éditeurs, pages 33–40, Edmonton, Alberta, Canada, May 31 - June 1 2003.
- [Grover *et al.*, 2004] Claire Grover, Ben Hachey et Chris Korycinski. A Rhetorical Status Classifier for Legal Text Summarisation. In *Text Summarization Branches Out Workshop held in conjunction with ACL’2004*, pages 35–42, Barcelona, Spain, 25-26 July 2004.
- [Hahn, 1990] Udo Hahn. Topic Parsing : Accounting for Text Macro Structures in Full-Text Analysis. *Information Processing and Management*, 26(1) :135–170, 1990.
- [Hearst, 1997] Marti A. Hearst. TextTiling : Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1) :33–64, 1997.
- [Hepple, 2000] Mark Hepple. Independence and Commitment : Assumptions for Rapid Training and Execution of Rule-based Part-of-Speech Taggers. In *the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*, pages 278–285, October 2000.
- [Hovy et Lin, 1999] Eduard Hovy et Chin Yew Lin. Automated Text Summarization in SUMMARIST. In *Advances in Automatic Text Summarization*, Inderjeet Mani et Maybury Mark T., éditeurs, pages 81–94. The MIT Press, 1999.
- [Jean Carletta et Anderson, 1997] Stephen Isard Jacqueline C. Kowtko Gwyneth Doherty-Sneddon Jean Carletta, Amy Isard et Anne H. Anderson. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 1(23) :13–31, 1997.
- [Jing et McKeown, 1999] Hongyan Jing et Kathleen R. McKeown. The Decomposition of Human-Written Summary Sentences. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, M. Hearst, Gey. F. et R. Tong, éditeurs, pages 129–136, University of California, Beekely, August 1999.
- [Jing, 2002] Hongyan Jing. Using Hidden Markov Modeling to Decompose Human-Written Summaries. *Computational Linguistics*, 28(4) :527–543, 2002.
- [Kupiec *et al.*, 1995] Julian Kupiec, Jan O. Pedersen et Francine Chen. A Trainable Document Summarizer. In *SIGIR95*, pages 68–73, 1995. <ftp://parcftp.xerox.com/pub/hearst/kupiec-sigir95.ps>.

- [Lehman, 1995] Abderrafih Lehman. *Le résumé automatique à fragments indicateurs : RAFI*. PhD thesis, Université de Nancy-II, Nancy, France, 1995.
- [Lehman, 1997] Abderrafih Lehman. Automatic summarization on the web. A system for summarizing using indicating fragments. In *Computer-Assisted Information Searching on Internet Conference (RIA0'97)*, pages 112–122, Québec, Canada, 1997.
- [Lin et Hovy, 1997] Chin-Yew Lin et Eduard Hovy. Identifying Topics by Position. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, pages 283–290. Association for Computational Linguistics, March 31 - April 3 1997.
- [Lin et Hovy, 2003] Chin-Yew Lin et Eduard Hovy. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In *Proceedings of 2003 Language Technology Conference (HLT-NAACL 2003)*, pages 150–157, Edmonton, Canada, May 27 - June 1 2003.
- [Lin, 2001] Chin-Yew Lin. Summary Evaluation Environment, 2001. <http://www.isi.edu/~cyl/SEE>.
- [Lin, 2004] Chin Yew Lin. ROUGE : A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out Workshop held in conjunction with ACL'2004*, pages 74–81, Barcelona, Spain, 25–26 July 2004.
- [Luhn, 1959] H.P. Luhn. The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, pages 159–165, 1959.
- [Mailhot, 1996] Louise Mailhot. *Ecrire la décision : guide pratique de rédaction judiciaire*. Editions Yvon Blais, Québec, Canada, 1996.
- [Mailhot, 1998] Louise Mailhot. *Decisions, Decisions : a handbook for judicial writing*. Editions Yvon Blais, Québec, Canada, 1998.
- [Mani et al., 1998] Inderjeet Mani, David House, G. Klein, Lynette Hirshman, Leo Orbst, Thérèse Firmin, Michael Chrzanowski et Beth Sundheim. The TIPSTER SUMMAC Text Summarization Evaluation. Rapport Technique MTR 98W0000138, The Mitre Corporation, McLean, Virginia, 1998.
- [Mani et al., 2002] Inderjeet Mani, Thérèse Firmin, David House, Gary Klein, Beth Sundheim et Lynette Hirschman. *The TIPSTER SUMMAC Text Summarization Evaluation*, volume 8 de 8. Cambridge University Press, 2002.
- [Mani et Bloedorn, 1998] Inderjeet Mani et Eric Bloedorn. Machine Learning of Generic and User-Focused Summarization. In *the Fifteenth National Conference on Artificial Intelligence (AAAI'98)*, pages 821–826, 1998.
- [Mani et Maybury, 1999] Inderjeet Mani et Mark Maybury. *Advances in automatic text summarization*. Kluwer Academic Publishers, Boston, U.S.A., 1999.
- [Mani, 2001] Inderjeet Mani. *Automatic Text Summarization*. John Benjamins Publishing Company, 2001.
- [Mann et Thompson, 1987] William C. Mann et Sandra A. Thompson. Rhetorical Structure Theory : A Theory of Text Organization. Rapport technique, University of Southern California/ISI, Marina del Ray, CA, U.S.A., 1987.
- [Marcu, 1997] Daniel Marcu. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. PhD thesis, University of Toronto, 1997.

- [Marcu, 2000] Daniel Marcu. *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge/London, 2000.
- [Minel *et al.*, 2001] Jean-Luc Minel, Jean-Pierre Desclés, Emmanuel Cartier, Gustavo Crispino, Slim Ben Hazez et Agata Jackiewicz. Résumé automatique par filtrage sémantique d'informations dans des textes. *Revue Technique et Science Informatiques*, (3), 2001.
- [Minel, 2002] Jean-Luc Minel. *Filtrage sémantique : du résumé automatique à la fouille de textes*. Editions Hermès, Paris, France, 2002.
- [Moens *et al.*, 1996] M.-F. Moens, R. Gebruers et C. Uyttendaele. SALOMON : Final Report. Rapport technique, Katholieke Universiteit Leuven, 1996.
- [Mourad, 2000] Ghassan Mourad. Présentation de connaissances linguistiques pour le repérage et l'extraction de citations. In *RECITAL'2000*, Lausanne, 16-18 October 2000.
- [Mourad, 2001] Ghassan Mourad. *Analyse informatique des signes typographiques pour la segmentation de textes et l'extraction automatique des citations. Réalisation des Applications informatiques : SegATex et CitaRE*. PhD thesis, Paris-Sorbonne (Paris IV), Paris, France, November 2001.
- [Ono *et al.*, 1994] Kenji Ono, Kazuo Sumita et Seiji Miike. Abstract Generation Based on Rhetorical Structure Extraction. In *Proceedings of the International Conference on Computational Linguistics*, pages 344–348, Kyoto, Japon, 1994.
- [Over et Yen, 2004] paul Over et James Yen. An Introduction to DUC 2004 Intrinsic Evaluation of Generic New Text Summarization Systems. In *DUC03 : NAACL'2003 Workshop in Automatic Text Summarization*, pages 1–62. National Institute of Standards and Technology, 2004.
- [Paice, 1981] Chris D. Paice. The Automatic Generation of Literary Abstracts : An Approach Based on Identification of Self-Indicating Phrases. In *Information Retrieval Research*, O. R. Norman, S. E. Robertson, C. J. van Rijsbergen et P. W. Williams, éditeurs, London : Butterworth, 1981.
- [Papineni et Zhu, 2001] T. Ward Papineni, S. Roukos et W. Zhu. Bleu : a method for automatic evaluation of machine translation. Technical report rc22176 (w0 109-022), IBM Research Division, 2001.
- [Pastra et Saggion, 2003] Katerina Pastra et Horacio Saggion. Colouring Summaries BLEU. In *Proceedings of the European Association of Computational Linguistics workshop on "Evaluation initiatives in NLP : are evaluation methods, metrics and resources reusable ?"*, *EACL 2003*, pages 32–39, Hungary, Budapest, 2003.
- [Quinlan, 1989] J.R. Quinlan. Induction of Decision Trees. *Machine Learning*, pages 81–106, 1989.
- [Radev *et al.*, 2002a] Dragomir Radev, Eduard Hovy et Kathleen McKeown. Introduction to the Special Issue on Summarization. *Computational Linguistics*, 28(4) :399–408, 2002.
- [Radev *et al.*, 2002b] Dragomir R. Radev, Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Arda Çelebi, Qi Hong, Elliott Drabek et Dany Liu. Evaluation of Text Summarization in a Cross-lingual Information Retrieval Framework. Rapport technique, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, 2002. <http://tangra.si.umich.edu/~radev/papers/report.pdf>.

- [Radev *et al.*, 2002c] Dragomir R. Radev, Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Arda Çelebi, Hong Qi, Elliott Drabek et Danyu Liu. Evaluation of Text Summarization in a Cross-lingual Information Retrieval Framework. Rapport technique, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD, June 2002.
- [Radev *et al.*, 2003] Dragomir Radev, Jahna Otterbacher, Hong Qi et Daniel Tam. MEAD ReDUCs : Michigan at DUC 2003. In *DUC03*, pages 160–167, Edmonton, Alberta, Canada, May 31 - June 1 2003. Association for Computational Linguistics.
- [Saggion et Lapalme, 2000] Horacio Saggion et Guy Lapalme. Concept Identification and Presentation in the Context of Technical Text Summarization. In *ANLP/NAACL00-WS*, Udo Hahn, Chin-Yew Lin, Inderjeet Mani et Dragomir R. Radev, éditeurs, Seattle, WA, USA, April 2000. Association for Computational Linguistics.
- [Saggion et Lapalme, 2002] Horacio Saggion et Guy Lapalme. Generating Indicative-Informative Summaries with SumUM. *Computational Linguistics*, 28(4) :497–526, 2002.
- [Saggion, 2001] Horacio Saggion. *Génération automatique de résumés par analyse sélective*. PhD thesis, Université de Montréal, 2001.
- [Salton, 1988] Gerard Salton. *Automatic Text Processing*. Addison-Wesley Publishing Company, 1988.
- [Silber et McCoy, 2002] Gregory H. Silber et Kathleen McCoy. Efficiently Computed Lexical Chains As An Intermediate Representation in Automatic Text Summarization. *Computational Linguistics*, 28(4), 2002.
- [Smith *et al.*, 1995] J. C. Smith, Daphne Gelbart, Keith MacCrimmon, Bruce Atherton, John McClean, Michelle Shinehoft et Lincoln Quintana. Artificial Intelligence and Legal Discourse : The Flexlaw Legal Text Management System. *Artificial Intelligence and Law*, 3(1-2) :55–95, 1995.
- [Smith et Deedman, 1987] J. C. Smith et Cal Deedman. The Application of Expert Systems Technology to Case-Based Law. *ICAIL*, pages 84–93, 1987.
- [Spark-Jones et Galliers, 1995] Karen Spark-Jones et Julia R. Galliers. *Evaluating Natural Language Processing Systems : An Analysis and Review*. Numéro 1083 dans Lecture Notes in Artificial Intelligence. Springer, 1995.
- [Spark-Jones, 1972] Karen Spark-Jones. Exhaustivity and specificity. *Journal of Documentation*, 28 :11–21, 1972.
- [Spark Jones, 1999] K. Spark Jones. Automatic Summarizing : Factors and Directions. In *Advances in Automatic Text Summarization*, I. Mani et M. Maybury, éditeurs, pages 1–14. MIT Press, Cambridge MA, 1999.
- [Teufel et Moens, 1997] Simone Teufel et Marc Moens. Sentence Extraction as a Classification Task. In *Proceedings of the Workshop on Intelligent Scalable Text Summarization, ACL/EACL*, 1997.
- [Teufel et Moens, 2002] Simone Teufel et Marc Moens. Summarising Scientific Articles - Experiments with Relevance and Rhetorical Status. *Computational Linguistics*, 28(4) :409–445, 2002.

- [Teufel, 1998] Simone Teufel. Meta-Discourse Markers and Problem-Structuring in Scientific Texts. In *Proceedings of the Workshop on Discourse Relations and Discourse Markers at the 17th International Conference on Computational Linguistics*, M. Stede, L. Wanner et Eduard Hovy, éditeurs, pages 43–49, August 15 1998.
- [Uyttendaele *et al.*, 1996] C. Uyttendaele, M.-F. Moens et J. Dumortier. SALOMON : Abstracting of Legal Cases for Effective Access to Court Decisions. In *Proceedings of JURIX 96 Ninth International Conference on Legal Knowledge Based Systems*, pages 47–58. Tilburg : University Press, 1996.

Annexe A

Exemple d'une évaluation extrinsèque

Pour l'évaluation extrinsèque, basée sur le jugement humain, nous avons fait évaluer 120 résumés des jugements. Le corpus d'évaluation était composé de dix jugements de la Cour fédérale du Canada tirés au hasard parmi les jugements des années 2000 à 2004. Ces jugements ne faisaient pas partie du corpus d'entraînement.

Nous avons généré des résumés pour les dix jugements avec le cinq systèmes de production du résumé suivants :

1. *LetSum* notre système de production du résumé de textes juridiques,
2. *StartEnd* notre méthode simplifiée de résumé basée sur la position des segments dans les textes juridiques,
3. système de recherche *MEAD* [Radev *et al.*, 2002c] [Radev *et al.*, 2003],
4. Système de recherche et commercial français *Pertinence Mining*[Lehman, 1995] [Lehman, 1997],
5. Système commercial de *Microsoft Word* (option de résumé dans MS Word).

Nous avons ajouté les dix résumés rédigés manuellement par les résumeurs professionnels de la Cour fédérale. Nous avons fait évaluer chaque résumé par deux juristes différents. Nous avons uniformisé les formats de tous les résumés pour uniformiser l'évaluation. Pour *LetSum*, nous n'avons utilisé que le texte du résumé sans le format tabulaire indiquant les thèmes des extraits.

Nous avons divisé les 120 jugements en 12 paquets pour chacun des 12 avocats de l'équipe LexUM.

Chaque juriste a reçu une trousse d'évaluation, mais il n'était pas s'informer sur les méthodes de productions des résumés. Ils ont reçu les instructions présentées à la fin de cette annexe. L'évaluation porte sur deux parties : répondre à sept questions

R-Méthodes	Q1	Q2	Q3	Q4	Q5	Q6	Q7	L	C	P	A
1. LetSum	oui	c	tb	t p	5,0						
2. Per.Mining	non	oui	oui	oui	oui	oui	oui	pc	m	ap	4,0
3. StadtEnd	oui	c	tb	tp	5,0						
4. Humain	oui	c	tb	tp	5,0						
5. MsWord	non	oui	oui	oui	non	non	oui	ac	b	ap	3,5
6. LetSum	non	oui	oui	non	non	oui	oui	c	tb	ap	4,0
7. MEAD	non	oui	oui	non	non	non	non	pc	m	pp	2,5
8. Humain	non	oui	oui	non	non	oui	oui	ac	m	ap	3,0
9. MEAD	non	oui	oui	oui	oui	oui	oui	ac	b	ap	2,5
10. Per.Mining	oui	oui	oui	non	non	non	non	pc	m	pp	2,0

TAB. 9.1 – Évaluation d’un juriste sur les résumés de la trousse d’évaluation

juridiques sur les contenus des résumés, et indiquer le degré de : lisibilité (L), cohérence (C), pertinence des phrases (P) et acceptabilité (A) des résumés.

Le temps d’évaluation a été limité à une heure, mais plupart des participants ont fini l’évaluation en 30 à 45 minutes. La discussion de la fin de l’évaluation a été très encourageante : les avocats considèrent la production de résumé automatique comme un effort intéressant pouvant beaucoup aider les juristes dans leurs tâches de traitement des jurisprudences. Certains avocats disaient que certains résumés étaient très bien écrits par des humains alors qu’ils avaient été générés automatiquement. D’autres croyaient qu’ils n’y avait pas de résumés manuels alors que tous en avaient reçu deux. Cette expérience montre que, même pour les experts de domaine juridique, les résumés humains ne sont pas parfaits et qu’ils trouvent très utiles les résumés construits automatiquement.

Selon les conseils des juristes, mais aussi dans le guide de jurisprudence de la juge Mailhot [Mailhot, 1998], le thème RAISONNEMENT JURIDIQUE le plus important. Notre évaluation extrinsèque montre toutefois que les thèmes les plus appropriés pour le résumé sont INTRODUCTION et CONCLUSION.

Dans les pages suivantes, nous montrons une des trousse d’évaluation contenant les instructions, les questions juridiques et qualitatives et dix résumés.

La table 9.1 montre les résultats donnés par un des juristes. Nous indiquons aussi le système qui a produit le résumé, information que n’avait pas donnée au moment de l’évaluation.

Évaluation

Groupe

23 Septembre 2004

Centre de recherche en droit public
de
l'Université de Montréal

Évaluation de systèmes de résumés automatiques

Instruction

Notre recherche est concentrée sur la production de résumé automatique. Nous travaillons sur le domaine juridique comme une application de résumé automatique. Afin de produire un condensé, le système extrait automatiquement les informations à partir des documents sources et il les présente à l'utilisateur. Les documents source sont les jugements rendus par la Cour Fédérale du Canada.

Évaluation

Nous étudions la qualité des résumés générés automatiquement. Dans cette évaluation, nous considérons deux aspects : d'abord le contenu des résumés sur les idées clés des jugements et par la suite l'acceptabilité de résumé.

Les documents sources sont des jugements de la Cour Fédérale et les résumés ont été générés soit par différents systèmes soit par des résumeurs professionnels.

Vous allez recevoir un formulaire qui contient des résumés et des questions à répondre. Vous ne seriez pas informé sur les méthodes de production des résumés.

Pour évaluer chaque résumé deux dimensions sont prises en compte : contenu et qualité. Première partie de l'expérience porte sur les réponses aux questions spécifiques sur les documents. Deuxième partie de l'expérience examine la qualité des textes des résumés: la lisibilité, la cohérence et la pertinence. Il y a aussi une valeur d'acceptabilité sur la qualité générale du résumé.

Première partie, évaluation sur le **contenu** du résumé :

Pour chaque résumé, vous avez 7 questions. Si un lecteur (ou lectrice) peut déterminer les points en questions en lisant le résumé, ça veut dire que le résumé contient suffisamment d'informations pour couvrir les idées essentielles d'un jugement.

Nous vous demandons de lire les résumés et répondre oui ou non aux questions.

Deuxième partie, évaluation sur la **qualité** du résumé :
Afin d'évaluer les textes, on vous demande d'attribuer des valeurs qualitatives pour chaque résumé sur les critères suivants:

1. Lisibilité du résumé

Ce critère donne une appréciation globale du résumé. On évalue si le résumé est :

- Clair
- Assez clair
- Peu clair
- Incompréhensible

2. Cohérence du résumé

Ce critère concerne le fil conducteur du texte pour en assurer la continuité et la progression de l'information. On évalue si le résumé est :

- Très bonne
- Bonne
- Médiocre
- Très mauvaise

3. Pertinence des phrases du résumé

La pertinence des phrases est l'élément qui se rattache à ce que l'on veut démontrer.

Ceci mesure si les phrases du résumé contiennent un lien clair et direct avec le sujet dont il est question. On évalue si le résumé est :

- Très pertinent
- Assez pertinent
- Peu pertinent
- Non pertinent

Nous vous demandons d'attribuer une valeur d'acceptabilité sur la qualité générale du texte de résumé. Cette valeur est entre 0 et 5 pour chaque résumé (0 pour un résumé inacceptable et 5 pour un texte acceptable). Les résumés avec valeur 2.5 jusqu'à 5 sont considérés acceptables.

Cette expérience ne mesure que la qualité des textes et non vous comme lecteur, d'ailleurs les réponses sont anonymes. Les résultats de cette expérimentation seront utilisés afin de mesurer la valeur de notre système de résumé par rapport à d'autres méthodes.

Nous vous remercions de votre participation à cette évaluation.
Votre jugement est très apprécié et qui nous aidera dans notre recherche.

Atefeh Farzindar
farzinda@iro.umontreal.ca

Évaluation

Ref :

A) Première partie, évaluation sur le contenu du résumé

Après avoir lu le résumé peut-on déterminer

1. Qui sont les parties en litige? Oui ----- Non -----

2. Quel est le problème en litige? Oui ----- Non -----

3. Les questions de droit soulevées (issues)? Oui ----- Non -----

4. Comment le juge a appliqué le droit aux faits? Oui ----- Non -----

5. Les motifs couvrent-ils les questions de droit? Oui ----- Non -----

6. Le résumé contient-il les motifs déterminants pour arriver à la conclusion? Oui ---- Non ---

7. Le résultat final de la cour? Oui ----- Non -----

Commentaires -----

Ref :

B) Deuxième partie, évaluation sur la qualité du résumé

I) Choisissez une valeur d'acceptabilité pour les critères suivants:

1. Lisibilité du résumé

Clair -----

Assez clair -----

Peu clair -----

Incompréhensible -----

2. Cohérence du résumé

Très bonne -----

Bonne -----

Médiocre -----

Très mauvaise -----

3. Pertinence des phrases du résumé

Très pertinent -----

Assez pertinent -----

Peu pertinent -----

Non pertinent -----

II) Attribuez une valeur d'acceptabilité sur la qualité générale du résumé (0 pour un résumé inacceptable et 5 pour un texte acceptable) -----

Commentaires -----

Les résumés à évaluer

[1] This is an application for judicial review of a decision of a delegate of the Minister of Citizenship and Immigration (the "Minister") dated May 6, 2002, wherein the Minister determined that the Applicant constitutes a danger to the public in Canada pursuant to s.53(1) of the Immigration Act. He was found to be a stateless person by the Canadian government. In 1993 he was declared a Convention refugee and came to Canada. He received permanent residence in 1995. [3] On November 7, 2000, the Applicant was convicted of trafficking cocaine, contrary to s.5(1) of the Controlled Drugs and Substances Act, S.C. 1996, c.19. [4] On June 5, 2002, the Applicant received notice of the intention of Citizenship and Immigration Canada to seek an opinion from the Minister declaring that the Applicant is a danger to the public in Canada. [6] The Minister's decision In forming my opinion, I have considered the Ministerial Opinion Report and the documentary evidence presented by local Immigration officials to support their recommendation that Tam Thanh TRAN, born June 23, 1976, citizen of Vietnam, constitutes a danger to the public pursuant to subsections 70(5) and 53(1) of the Immigration Act , as well as the information contained in the Request for Minister's Opinion report dated January 22, 2002 and supporting material. [13] I agree with the argument advanced by counsel for the Applicant that the statements relied upon by the Respondent still leave a primary question outstanding: Why was the decision made? [14] The definition of reasons, considered in the present context, requires that the totality of the evidence with respect to the Applicant be employed in an analytical argument justifying the danger decision reached. I cannot find that the cryptic contentious statements in the ministerial reports so quality. [15] As a result, I find that the Minister's decision was reached in error of law.

You were assessed in the occupation of Art Director with reference to the National Occupational Classification (NOC) 5241, which sets out the applicable employment requirements for Canada and the main duties for each occupation.

The Applicant says that by adopting this narrow interpretation, the Visa Officer had imposed an additional requirement into the NOC 1122.2 definition, thereby committing an error in law.

[19] The Respondent also submits the Visa Officer was correct in his interpretation of the occupational definition of NOC 1122.2 and that he did not introduce any additional requirements into this definition.

[27] The Applicant relies on the decision in *Uy*, supra, in support of his argument that a formal assessment, including the award of points for the identified factors, must be carried out in respect of his intended occupation.

I do not understand this jurisprudence to require a visa officer to continue an assessment in a given occupational category after it has become clear that the applicant cannot obtain the required number of points to be granted landing.

[31] However, it is clear that the Visa Officer based his decision not to conduct an assessment on his determination that the Applicant lacked the employment requirements for the position.

it may be an error to state that "is usually required" means that the educational requirement must be met, except where there are significant and substantial factors that persuade the visa officer that the occupational requirements can be overcome.

Mr. Charles Webster (the "Applicant") brings this application for judicial review pursuant to the Federal Court Act R.S.C. 1985 c. F-7 as amended section 18.1. The Applicant seeks a declaration injunction and writ of prohibition enjoining the Minister of National Revenue (the "Minister") from taking any of the actions set out in sections 225.1(1)(a) to (g) of the Income Tax Act R.S.C. 1985 (5th Supp.) c. 1 as amended (the "Act"). The Applicant is a chartered accountant and a Canadian taxpayer. In his 1991 income tax return he claimed an addition to his "cumulative Canadian exploration expense" account as defined in section 66.1(6)(a) of the Act relative to an investment in an oil and gas exploration and development business. In his income tax returns for 1991 1992 and 1993 he claimed certain deductions relative to the cumulative Canadian exploration expense account including in 1992 and 1993 deductions for interest charged on a promissory note used to finance his investment and further deductions for general and administrative expenses related to the operation of the oil and gas business. In 1995 the Minister issued Notices of Reassessment dated April 11 1995 in which he was refused the deductions referred to above except for some minor amounts claimed for general and administrative expenses in the 1992 and 1993 income tax returns. The Applicant objected to the reassessments by Notices of Objection dated June 30 1995. In that agreement he agreed to be bound by the decision in another case. The similarity of the present case with the "test" case was a matter to be agreed upon by the Applicant and the Minister. I agree with the Respondent that the applicable test of "similarity" is a subjective one; otherwise the Court would be reviewing an agreement between the parties and that is not contemplated by section 225.1(5) in my opinion. The application is dismissed with costs to the Respondent.

Application for judicial review of decision of Minister of Indian Affairs and Northern Development withholding approval of licence issued by Nunavut Water Board (NWB) to Municipality of Iqaluit. In 1993, Government of Canada entered into Agreement between the Inuit of the Nunavut Settlement Area and her Majesty the Queen in the Right of Canada. Inuit of Nunavut exchanged Aboriginal title to lands and waters in Nunavut Settlement Area for treaty rights set out in Agreement. In September 2000, Municipality applied to NWB for renewal of water licence to expire on December 31, 2000. Minister not willing to approve licence as issued to Municipality. Whether Agreement provides, expressly or by implication, licensing decisions of NWB subject to Minister's approval. Under Agreement, negotiating parties intended regulatory regime set out in Northern Inland Waters Act (NIWA) to continue except as varied by Agreement. No dispute water use licences issued by Northwest Territories Water Board under NIWA, Northwest Territories Waters Act (NTWA) required ministerial approval. Nunavut Waters and Nunavut Surface Rights Tribunal Act also including ministerial approval requirement. Whether, after creation of NWB and before coming into force of Nunavut Waters Act, provisions of Agreement operate to remove requirement for ministerial approval. Agreement not intended to be complete source of power, authority for NWB. NTWA continued to have force and effect in Nunavut, before and after ratification of Agreement, creation of NWB, creation of Nunavut on April 1, 1999. On face of Agreement, art. 13.7.1 not providing for ministerial approval. Art. 13 not intending to describe in detail regulatory regime for water management. Requirement for ministerial approval set out in NIWA not removed by Agreement, continues as provided for in NIWA. Other provisions of Agreement explicitly providing for ministerial approval. Goal of Agreement to ensure meaningful role for Inuit in decision-making. Requirement for ministerial approval not resulting in inconsistency, contradiction between provisions of Agreement, applicable legislation. Minister had legal authority to withhold approval of licence issued by NWB. Application dismissed. Northern Inland Waters Act, R.S.C., 1985, c. N-25. Northwest Territories Waters Act, S.C. 1992, c. 39. Nunavut Waters and Nunavut Surface Rights Tribunal Act, S.C. 2002, c. 10.

[1] This is a motion by the Defendant in a patent infringement action for an order striking out paragraph 25 of the Plaintiff's amended Statement of Claim or, alternatively, for an Order requiring the Plaintiff to provide further particulars with respect to said paragraph as well as with respect to paragraph 24 of Plaintiff's amended Statement of Claim. [2] The Defendant is the inventor and owner of Canadian Patent No. 2,219,072 (the 072 Patent), issued on August 22, 2000. [5] Paragraph 24 of the Plaintiff's amended Statement of Claim alleges that the 072 Patent is invalid for failing to comply with section 28.3 of the Patent Act, R.S.C., 1985, c. P-4 (the Act) in that the subject matter was obvious on the claim date to a person skilled in the art or science. [7] Paragraph 25 of the Plaintiff's amended Statement of Claim c) and admits that it is not relying at all on said paragraph. [13] Therefore, with respect to striking out paragraph 25 of the Plaintiff's amended Statement of Claim, I must side with the Plaintiff and conclude for the above reasons that it is not clear and obvious that said paragraph 25 should be struck out. Consequently, that portion of the Defendant's motion is denied. [14] I now turn to the Defendant's request for particulars regarding paragraph 25 of the Plaintiff's amended Statement of Claim, and in particular that the Plaintiff be requested, inter alia, to provide the basis on which the Plaintiff claims that the specification of the 072 Patent is insufficient because it fails to describe the best mode. [15] I agree with the Defendant that the Plaintiff should be ordered to answer that specific request. [17] As for the particulars requested with respect to paragraph 24 of the Plaintiff's amended Statement of Claim, I also agree with the Defendant that the Plaintiff should be ordered to indicate which portions of the documents listed in paragraph 24 of the Plaintiff's amended Statement of Claim are relied upon in support of the allegation that the 072 Patent is invalid for failing to comply with section 28.3 of the Act.

In fact, the plaintiff is seeking judicial review of a decision by the defendant dated April 28, 1994 pursuant to the Supplementary Retirement Benefits Act ("the Act"). In his application the plaintiff himself admitted that under the applicable rules he had until May 28, 1994 to file an application for judicial review. The defendant's arguments were that the Act should be applied as it stands and that it was the Act which applied at the time the plaintiff took his decisions to back to work. I can readily understand the extent of the plaintiff's frustration, finding himself in a legal dead end, when he had thought he would be improving his situation by returning to work for the Department of National Defence 20 years ago, and for the perverse consequences of that decision to be felt at the time of his retirement, the time when he most needed to have a reasonable pension fund, the time when it was most difficult for him to continue in the labour market as he was over 60 years old. At this stage, I do not have to decide the merits of the case, but rather the legitimacy of the motion, based on what it concerns, and in particular the existence of a serious issue to be argued in the Federal Court. The plaintiff also persuaded the Court that he had a serious issue to argue in the Federal Court and that it was in the interest of justice for him to be heard in order to assert his rights.

[2] It is settled that the Court will generally refuse to strike out "surplus statements" which are not prejudicial, and that doubt is to be resolved in favour of permitting the pleading so that the relevant evidence in support of the pleading may be brought before the Trial Judge; *Imperial Parking Ltd. v. Barrington* 1995 , 104 F.T.R. 132, F.C.J. No.1638 F.C.T.D. per Prothonotary Hargrave, and *Huzar v. Canada* 1997 , 139 F.T.R. 81, F.C.J. No.1556 F.C.T.D. , where Prothonotary Hargrave held at paragraphs 13, 14 and 15 : ... It must be plainly, obviously and beyond doubt that a pleading is futile and will not succeed before it will be struck out. [3] When it is alleged that an action is scandalous, frivolous, or vexatious, under Rule 419 1 c the test is as stringent as or even more stringent than that under Rule 419 1 a Rule 419 is the predecessor to Rule 221 ... When there are contentious or serious issues of law, disputed points of law, or uncertain points of law, they ought not be determined on a summary motion to strike out, but rather left for a decision at Trial when all the facts are known ... The final relevant point of procedure is that a Court will not strike out statements that are merely surplus, provided no prejudice flows from them ... In this case, the plaintiff submits that the prior proceedings are relevant, and that the Court in a 31-page Decision, considered and constructed the individual patents with respect to the plaintiff's formulation in issue, and that the Federal Court of Appeal on April 2, 2001 upheld this Decision and made a finding at paragraph 16: There can be no infringement of the patent claims if both properties are not present in the cefuroxime axetil that is utilized.

Case management conference to explain why eight-year-old action, on which nothing substantive happening since examinations for discovery in 1997, should not be dismissed for delay. Action challenging Minister's decision duty owed on importation of goods to Canada. Two approaches to dismissal for delay, want of prosecution. Classic test set out in (1990), 36 F.T.R. 77 (F.C.T.D.): whether inordinate delay, whether delay inexcusable, whether defendants likely to be seriously prejudiced by delay. Although decided before 1998 changes to Federal Court Rules, (1998), 152 F.T.R. 96 (F.C.T.D.) holding classic test applied under post-1998 Rules, reflecting r. 167 concept of dismissal for undue delay. Reasons given by defendant for delay: illness, both plaintiff's own and in family, and death in family. While inordinate delay, and delay not explained, only view of counsel for defendant as to difficulty in now finding witnesses. After such long delay, prejudice often presumed. Second approach more apt in case of lay litigant. Employed in , [1997] 1 W.L.R. 640 (H.L.): if lay litigant engaging in wholesale disregard for time limits provided in rules of court, breach should be looked at not only from viewpoint of prejudice to litigants, but also in light of abuse, prejudice to due administration of justice. In , [1998] 1 W.L.R. 1426 (C.A.), Court of Appeal of view such abuse of process separate ground for striking out action, not requiring defendant to demonstrate prejudice. Lack of reasonable plan to bringing action to conclusion indicating plaintiff having no real interest in doing so. Action, as conducted, constituting abuse. Unfair to defendant, to Court whose resources limited, to taxpayers who must pay Court's bills, to other litigants who look to Court for reasonably quick resolution of disputes. Action dismissed. Federal Court Rules, 1998, SOR/98-106, r. 167.

[1] I have reviewed the written submissions of both parties on costs. [2] FACTS On October 4, 2002, the plaintiff made an offer to settle to the defendants, pursuant to subsection 420 1 of the Federal Court Rules, 1998 Rules in the amount of Cdn 98,988.50 in full and final settlement of this matter in principal, interest and costs.

[4] 1 Unless otherwise ordered by the Court, where a plaintiff makes a written offer to settle that is not revoked, and obtains a judgment as favourable or more favourable than the terms of the offer to settle , the plaintiff shall be entitled to party-and-party costs to the date of service of the offer and double such costs, excluding disbursements, after that date.

[6] 13 In *Feherguard Products Ltd. v. Rocky's of B.C. Leisure Ltd.* , 1994 53 C.P.R. 3d 417, 1994 F.C.J. No. 2012, Stinson Taxing Officer stated that there is an automatic triggering effect of a valid offer to settle that does not require the Court's intervention former rule 344.1 being similar to current rule 420 : para 9 ... [7] emphasis added 16 In *Sanmammias Compania Maritima S.A. v. The "Netuno"* , 1995 102 F.T.R. 181, 1995 F.C.J. No. 1442, Tremblay-Lamer J. followed the reasoning of the "PRESTIGIOUS", supra , and decided that a doubling of costs would be excessive since the settlement offer was made only six days before trial, thereby awarding costs from the date of the offer at 150 . [8] 15 In these circumstances, if I were to issue special directions concerning costs after March 30, 2000, in view of the unaccepted offer to settle, it would be that costs after that date, excluding disbursements, be awarded at 1.33 units, or 133 percent, of the units for party-and-party costs prior to that date. [9] O R D E R 1 Therefore, the plaintiff should be entitled to party-and- party costs to the date of service of the settlement offer, i.e. October 4, 2002 and 150 of party-and-party costs thereafter, excluding disbursements, after that date.

Marcel Charlebois's affidavit gives a general description of the complaints drawn to the Commissioner's attention since the Act came into force in 1988. It also discusses the numerous occasions on which the Commissioner has intervened with Air Canada since that date, in particular in connection with complaints concerning the failure to provide French-language services at the airport in question. The affidavit also describes the systemic problem observed at that airport and lists the types of ground services concerned in the 26 separate complaints received since 1990.

Do those documents filed by the Commissioner in support of his motion that concern similar complaints and relate to all the ground services provided by Air Canada at the Halifax airport exceed the scope of the remedy provided for in the Act?

Air Canada's position is therefore that the Commissioner may only apply for a remedy limited to facts relating to a specific complaint, the investigation of that complaint and the resulting reports and recommendations. On the other hand, the following provision, section 79, states that information relating to any "similar" complaint in respect of "the same federal institution" may be admitted as evidence.

I therefore agree with the Commissioner's position that the remedy is not limited to certain types of ground services listed in Paul Comeau's two specific complaints but may apply to all ground services provided by Air Canada at the Halifax airport.

Nothing in the Act indicates that information in closed files, namely files already considered by the Commissioner, cannot be reconsidered in reviewing similar complaints in respect of the same federal institution.

Furthermore, subsection 64(2) authorizes the Commissioner to follow up on his recommendations where in his opinion adequate and appropriate action has not been taken thereon within a reasonable time.