



HAL
open science

Analyse Informatique de Guides de Bonnes Pratiques Cliniques

Gersende Georg

► **To cite this version:**

Gersende Georg. Analyse Informatique de Guides de Bonnes Pratiques Cliniques. Autre [cs.OH].
Université Pierre et Marie Curie - Paris VI, 2006. Français. NNT: . tel-00102511

HAL Id: tel-00102511

<https://theses.hal.science/tel-00102511>

Submitted on 1 Oct 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : INFORMATIQUE MÉDICALE

présentée par

M^{elle} Gersende GEORG

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PARIS 6

Sujet :

Analyse Informatique de Guides de Bonnes Pratiques Cliniques

Soutenue le 6 septembre 2006 devant le jury composé de :

Monsieur Alain-Jacques VALLERON	Université Paris 6	(Président)
Monsieur Bernard MOULIN	Université Laval, Québec, Canada	(Rapporteur)
Monsieur Domenico PISANELLI	CNR – ISTC, Italie	(Rapporteur)
Monsieur Gilles CHATELLIER	Université Paris 5	(Examineur)
Monsieur Patrice DOSQUET	Haute Autorité de Santé	(Examineur)
Madame Marie-Christine JAULENT	INSERM UMR_S 729	(Directrice de thèse)

Remerciements

Je remercie le Professeur Alain-Jacques Valleron de l'intérêt qu'il a bien voulu porter à mon travail en acceptant de participer à ce jury.

Le Docteur Patrice Dosquet a suivi les développements de cette recherche et a été le premier à m'encourager en ce qui concerne ses retombées potentielles. Je tiens à le remercier de ses encouragements.

Je remercie le Professeur Gilles Chatellier pour son soutien et la confiance qu'il m'a toujours témoignée. Ses conseils m'ont été précieux et m'ont enrichie tant au point de vue scientifique qu'humain. Je le remercie également de m'avoir offert l'opportunité de travailler sur le projet EsPeR.

Je remercie Marie-Christine Jaulent pour son écoute et sa compréhension pendant tout le déroulement de cette thèse. Au cours de ces années, j'ai pu apprécier son soutien, sa détermination et ses grandes qualités humaines.

Je suis heureuse que Domenico Pisanelli ait pu apporter le regard de son expérience de l'informatique médicale. Je le remercie d'avoir accepté de juger cette thèse.

Les travaux du Professeur Bernard Moulin ont servi de point de départ à l'approche utilisant les opérateurs déontiques. Je le remercie donc sincèrement d'avoir accepté de jouer le rôle de rapporteur, et ses commentaires m'ont beaucoup encouragée.

Je voudrais remercier le Docteur Isabelle Colombet pour sa gentillesse, son aide, et ses conseils. Le travail avec elle sur le projet EsPeR m'a beaucoup apporté.

Le Docteur Pierre Durieux s'est intéressé à ce travail, malgré son scepticisme envers les systèmes informatiques, et a accepté d'être un des premiers évaluateurs du prototype.

Je remercie également le Professeur Jean-Michel Halimi d'avoir accepté de faire partie des évaluateurs du prototype.

Je tiens à remercier le Professeur Joël Ménard pour son enseignement informel sur les Guides de Bonnes Pratiques, que ce soit au niveau de leur élaboration que de leurs retombées en pratique.

François Rastier m'a permis d'assister à ses séminaires « Sémantique des textes » me permettant de découvrir un domaine que je connaissais que très peu et pour lequel il m'a beaucoup appris et a attisé ma curiosité.

Toute l'équipe du SPIM et de l'Inserm UMR_S 729, et en particulier le Professeur Patrice Degoulet, m'a accueilli chaleureusement au sein du laboratoire.

Je tiens à remercier plus particulièrement quelques collègues pour avoir bien voulu relire différents chapitres de cette thèse : Julie Niès, qui a fait une relecture très riche en suggestions, ainsi que David Ouagne et Iulian Alecu.

Je remercie mes parents, pour m'avoir soutenue durant toutes ces années, et j'espère qu'ils trouveront la concrétisation de leur soutien dans cette thèse. Les compétences d'enseignante de ma mère ont également été mises à contribution lors des dernières corrections du manuscrit.

Je remercie Jacqueline d'avoir toujours été présente pour moi ces dernières années. Je voudrais également remercier mes grands-parents et plus particulièrement mon grand-père à qui je dédie cette thèse.

Franco Zaroli a été le premier enseignant à avoir cru à mon potentiel. J'espère qu'il trouvera dans cette thèse l'aboutissement du travail commencé à Strasbourg.

Je remercie de tout cœur Géraldine, mon amie depuis très longtemps ; nos conversations, parfois très longues, ont été une source d'énergie constante ... je la remercie aussi bien évidemment pour sa relecture et tutti quanti ...

Enfin, les mots ne sauraient exprimer le soutien exceptionnel de celui qui partage ma vie, Marc ...

Résumé :

Les Guides de Bonnes Pratiques Cliniques sont des documents normalisés destinés à l'amélioration de la qualité des soins. L'informatisation des Guides de Bonnes Pratiques est un enjeu majeur pour leur diffusion mais aussi pour les différentes étapes de leur cycle de vie depuis leur rédaction. De nombreux travaux de recherche ont été consacrés à l'informatisation des Guides de Bonnes Pratiques. Parmi eux, les approches documentaires utilisent les développements récents des langages à base de balises pour la structuration des documents, facilitant le traitement informatisé de ces documents. Nous proposons une extension de l'approche documentaire qui est basée sur une analyse intelligente des marqueurs de structuration des recommandations utilisant un analyseur à états finis. Cette approche a été intégrée à l'environnement informatique G-DEE (*Guideline Document Engineering Environment*) dédié à l'étude des Guides de Bonnes Pratiques, et peut servir pour une aide à la rédaction, un support au consensus lors de l'écriture, et un support pour la maintenance et l'actualisation des Guides de Bonnes Pratiques.

Mots-clés :

Guides de Bonnes Pratiques ; ingénierie documentaire ; opérateurs déontiques ; automates à états finis ; G-DEE ; langages de balisage.

Title: Computer Analysis of Clinical Guidelines

Abstract:

Clinical Guidelines are standardized documents developed to improve the quality of medical care. The computerization of clinical guidelines has attracted much interest in recent years, as it would not only facilitate their dissemination but also improve the knowledge-based process through which they are produced. Document-based approaches are derived from document engineering techniques using markup languages to produce structured documents. We propose to extend the document-based approach by re-introducing some degree of automatic processing, dedicated to the recognition of linguistic markers using finite-state transition networks. This approach has been integrated in our G-DEE (*Guideline Document Engineering Environment*) software environment, which is dedicated to the study of Clinical Guidelines but can also be extended to support the development and maintenance of Clinical Guidelines.

Keywords:

Clinical Guidelines; deontic operators; finite-state transition networks; G-DEE; document engineering; markup languages.

Intitulé et adresse de l'unité où la thèse a été préparée :



Santé Publique et Informatique Médicale (SPIM) ; UMR_S 729
Centre des Cordeliers
15, rue de l'École de Médecine
75006 Paris, FRANCE

Contenu de la thèse

CHAPITRE 1

INTRODUCTION	1
1.1 Objectif de la thèse et démarche suivie	2
1.2 Présentation du document	4

CHAPITRE 2

LES GUIDES DE BONNES PRATIQUES, DE LEUR ÉLABORATION À LEUR INFORMATISATION	7
2.1 Cycle de vie des Guides de Bonnes Pratiques	8
2.2 Élaboration des Guides de Bonnes Pratiques	9
2.3 Rôle des Guides de Bonnes Pratiques	10
2.4 Obstacles à l'utilisation des Guides de Bonnes Pratiques	12
2.5 Informatisation des Guides de Bonnes Pratiques : approches centrées sur les connaissances	16
2.5.1 La Syntaxe Arden et les MLM	18
2.5.2 Le Format d'échange de Guides de Bonnes Pratiques (<i>Guideline Interchange Format</i> ou GLIF)	20
2.5.3 <i>PROforma</i>	24
2.5.4 GUIDE	28
2.5.5 Aperçu des autres approches	31
2.6 Approches documentaires	41
2.6.1 Hypertext Guideline Markup Language	41
2.6.2 Guideline Elements Model (GEM)	42
2.7 Interfaces dédiées aux approches documentaires	45
2.7.1 GEM-Cutter	45
2.7.2 Guide-X / Stepper	46
2.7.3 GMT (Guideline Markup Tool) / DELT/A	48
2.7.4 DeGeL / URUZ	49
2.8 Conclusion	50

CHAPITRE 3

CARACTÉRISATION LINGUISTIQUE DES GUIDES DE BONNES PRATIQUES	53
3.1 Opérateurs déontiques dans les Guides de Bonnes Pratiques	54
3.1.1 Verbes déontiques dans la littérature	54
3.1.2 Méthodologie pour l'identification des opérateurs déontiques	55
3.2 Distributions des opérateurs déontiques dans les textes	60
3.3 Conclusion	64

CHAPITRE 4

APPROCHES LINGUISTIQUES	65
4.1 Caractérisation de l'approche	66
4.2 Les techniques du traitement du document	67
4.2.1 Le marquage sémantique des balises	68
4.2.2 Le marquage sémantique du contenu	71
4.3 Les techniques de l'Extraction d'Informations	73

CHAPITRE 5	
L'ENVIRONNEMENT G-DEE	77
5.1 Description générale de G-DEE	78
5.2 Les zones d'affichage de G-DEE	79
5.2.1 Analyse et structuration du texte	79
5.2.2 Structuration selon le modèle GEM	80
5.2.3 Génération de règles de décision	80
5.3 Aperçu des fonctionnalités	81
5.3.1 Analyse d'une sélection ou du texte entier	81
5.3.2 Affichage <i>Front-Scope</i> et/ou <i>Back-Scope</i>	81
5.3.3 Modification / Validation d'une phrase	84

CHAPITRE 6	
TECHNIQUES D'ANALYSE LINGUISTIQUE	87
6.1 Stratégie de reconnaissance de motifs	88
6.1.1 Reconnaissance des opérateurs déontiques	88
6.1.2 Reconnaissance des <i>scopes</i>	90
6.2 Règles de description syntaxique	93
6.2.1 Composition du corpus	93
6.2.2 Opérateurs déontiques en contexte	93
6.2.3 Forme monadique ou dyadique des opérateurs déontiques	101
6.2.4 Voix active ou passive des opérateurs déontiques	101
6.2.5 Forme singulière et plurielle des opérateurs déontiques	102
6.2.6 Formes infinitives	103
6.2.7 Les connecteurs : marqueurs du contenu des <i>scopes</i>	104
6.3 Stratégie de pré-traitement pour l'analyse de « textes réels »	106
6.3.1 Pré-traitements et Guides de Bonnes Pratiques	107
6.3.2 Nécessité d'un pré-traitement	108
6.3.3 Règles d'usage de la ponctuation	108
6.3.4 Pré-traitements des phrases contenant des signes de ponctuation	115
6.3.5 Pré-traitements des phrases sans signes de ponctuation	121

CHAPITRE 7	
FONCTIONNEMENT DE L'ANALYSEUR & INTÉGRATION DANS G-DEE	123
7.1 Fonctionnement de l'analyseur	124
7.1.1 Étapes préliminaires	124
7.1.2 Balisage des opérateurs déontiques	127
7.1.3 Balisage des <i>scopes</i>	130
7.1.4 Balisage des formes infinitives	131
7.1.5 Balisage des connecteurs de condition	134
7.2 Performance et Complexité de l'algorithme	137
7.3 Intégration des fonctionnalités	138
7.3.1 Règles de transformation	139
7.3.2 Structure du fichier XML du texte balisé	141
7.3.3 Interprétation de surface du texte balisé	141

CHAPITRE 8	
COUVERTURE SYNTAXIQUE ET TRAITEMENT DE FORMES PARTICULIÈRES	151
8.1 Préambule	152
8.2 Pré-traitement et « phrases réelles »	152
8.3 Formes adjectivales	154
8.4 Formes particulières utilisant les auxiliaires et les verbes déontiques	155

CHAPITRE 9	
PERFORMANCES ET ÉVALUATION PRÉLIMINAIRE DE G-DEE	157
9.1 Évaluation préliminaire par un jeu de test	158
9.2 Évaluation par des experts	159
9.2.1 Protocole d'évaluation	160
9.2.2 Étude de la concordance entre experts	160
9.2.3 Étude des accords entre experts	162
9.2.4 Étude des désaccords entre experts	162
9.2.5 Performance « brute » de G-DEE	162
9.2.6 Les FN détectées par des experts	164
9.2.7 Les FP détectées par des experts	166
9.3 Conclusion	168
CHAPITRE 10	
CONCLUSION	169
10.1 Préambule	170
10.2 Impact potentiel de G-DEE sur le cycle de vie des Guides de Bonnes Pratiques	171
10.3 Perspectives pour l'élaboration des Guides de Bonnes Pratiques	172
10.3.1 G-DEE pour l'aide à la rédaction	172
10.3.2 G-DEE pour l'aide à la maintenance et à l'actualisation	175
10.4 Perspectives pour l'informatisation des Guides de Bonnes Pratiques	176
10.5 Conclusion générale	177
RÉFÉRENCES	179
ANNEXES	193
Annexe 1 – Publications réalisées durant la thèse	194
Annexe 2 – Guides de Bonnes Pratiques utilisés pour l'analyse linguistique	195
Annexe 3 – Guides de Bonnes Pratiques utilisés pour établir la grammaire	198
Annexe 4 – Exemple d'un Guide de Bonnes Pratiques analysé par G-DEE	200
Annexe 5 – Résultats d'évaluation des experts avec leur appréciation pour le Guide de Bonnes Pratiques de l'AVC	209

Chapitre 1

Introduction

Le traitement des documents médicaux est devenu un domaine important de l'Informatique Médicale, en raison du rôle central joué par ceux-ci dans tous les domaines de la Médecine depuis la diffusion des connaissances jusqu'à la pratique clinique et la gestion des soins. Les documents médicaux se répartissent en genres précis qui ont chacun une identité forte : dossiers cliniques, comptes rendus d'hospitalisation ou d'intervention, correspondances, articles scientifiques. Certains textes ont de plus un caractère normatif dans la mesure où ils visent à encourager des attitudes cliniques efficaces, qu'on désigne habituellement par le terme de « Bonnes Pratiques Cliniques ». Ces textes normatifs appelés Guides de Bonnes Pratiques ont pour fonction de décrire des « conduites à tenir » basées sur le concept d'*Evidence-Based Medicine*, ou médecine basée sur les preuves. Au plan de l'expression, ils partagent certaines propriétés des textes juridiques et réglementaires (Moulin & Rousseau, 1990) même si leur style est généralement moins prescriptif. Chaque Guide de Bonnes Pratiques aborde un thème unique, par exemple la prise en charge de l'hypertension artérielle, ou le suivi du patient diabétique.

Cependant, le processus de rédaction de ces documents reste complexe du fait de leur production par des comités d'experts ; de plus, leur mise en œuvre en pratique et leur acceptation par les praticiens se heurtent à de multiples obstacles. En effet, les principales critiques qu'on peut leur faire concernent autant leur structure (ils sont souvent très longs à lire), que leur contenu, parfois sujet à interprétation (malgré l'objectif de normalisation affiché). L'informatisation de ces documents peut être une opportunité pour améliorer les différentes étapes du cycle de vie des Guides de Bonnes Pratiques et constitue donc un enjeu important pour leur diffusion et leur succès. C'est cet aspect qui est essentiellement développé dans cette thèse dédiée à l'approche informatique des Guides de Bonnes Pratiques. Cette thèse est l'occasion de poser certaines questions fondamentales sur la nature des Guides de Bonnes Pratiques et à partir de leur description de proposer des approches techniques qui soient le mieux à même de permettre leur analyse informatique. Nous entendons ici le terme d'analyse dans deux sens complémentaires. D'une part, nous nous intéressons à des méthodes informatiques qui permettent d'étudier les Guides de Bonnes Pratiques en tant que textes et en ce sens qui permettent de supporter une recherche sur les Guides de Bonnes Pratiques en tant que documents. D'autre part, et dans un sens plus classique en Informatique Médicale, l'analyse des documents correspond à leur traitement automatique par des méthodes d'ingénierie documentaire, enrichies de certaines fonctionnalités d'extraction de contenu linguistique. Ces méthodes visent à terme à

automatiser certaines approches comme la structuration et le balisage, dont on montrera qu'elles jouent un rôle important dans la diffusion électronique des documents et leur encodage dans des formats normalisés.

1.1 Objectif de la thèse et démarche suivie

L'objectif de cette thèse consiste à proposer des techniques d'analyse identifiant des propriétés caractéristiques des Guides de Bonnes Pratiques à partir de leur contenu. L'originalité de notre approche va consister à reconnaître des éléments de contenu linguistique pour structurer le document.

Pour ce faire, il faut dans un premier temps caractériser le texte d'un point de vue linguistique. Cette étape descriptive va permettre de proposer des méthodes adaptées pour repérer les marqueurs structurant le texte des Guides de Bonnes Pratiques. Nous proposons donc de structurer le document autour de marqueurs linguistiques spécifiques qui apparaissent de façon caractéristique dans les Guides de Bonnes Pratiques comme dans d'autres textes normatifs appartenant à d'autres domaines de spécialité. La reconnaissance des formes de surface de tels marqueurs permettra de structurer les portions du texte qu'ils encadrent grâce à une définition simple de la « Portée » de ces marqueurs. Il nous restera ensuite à donner une existence formelle à cette structuration en utilisant une technique de gestion documentaire bien connue : le balisage.

Pour cette reconnaissance de marqueurs, nous utiliserons des techniques empiriques de Traitement Automatique du Langage Naturel (TALN), et plus spécifiquement les techniques empruntées à l'Extraction d'Informations (Cowie & Wilks, 2000) (détaillées dans le chapitre 4) qui permettront de définir une « grammaire » des marqueurs de structure.

Nous proposons une architecture logicielle pour sous-tendre les différentes expériences d'analyse des Guides de Bonnes Pratiques et qui pourra servir de socle au développement ultérieur d'environnements de structuration et/ou d'encodage automatiques. Notre objectif consiste à créer un environnement documentaire qui prenne pour point de départ l'état de l'art des approches documentaires dans le contexte des Guides de Bonnes Pratiques. Des modèles tels que GEM (*Guideline Elements Model*) (Shiffman et al., 2000) ont proposé de structurer ces documents pour faciliter leur standardisation et leur informatisation. La difficulté de représenter formellement les connaissances médicales dans ces modèles a fait émerger le développement d'interfaces dédiées pour faciliter le passage du texte vers ces modèles informatiques. Néanmoins, ces éditeurs ne font que supporter un balisage manuel de ces documents. L'approche que nous proposons veut étendre l'état de l'art en y intégrant des éléments d'analyse automatique de contenu. Ces traitements linguistiques sont également basés sur une approche linguistique et descriptive des Guides de Bonnes Pratiques, qui apportent donc des éléments de réflexion plus fondamentaux sur ces documents.

1.1. Objectif de la thèse

Enfin, il nous faudra valider l'approche sur un prototype et sur un ensemble non trivial de données de Guides de Bonnes Pratiques qui permettra de mesurer autant la performance de l'analyseur intégré à G-DEE (*Guideline Document Engineering Environment*) (Georg & Jaulent, 2005) que de mesurer la validité de certaines hypothèses sous-jacentes. Des évaluations inspirées de celles effectuées pour les systèmes d'Extraction d'Informations permettront de juger de la performance de l'analyseur. Les experts humains permettront de valider les analyses de Guides de Bonnes Pratiques produites par le système G-DEE en comparant son identification de recommandations élémentaires à leur propre compréhension du document en contexte.

L'approche que nous venons de décrire doit donc articuler soigneusement la théorie et la pratique, et ne peut se réduire aux techniques empiriques d'analyse du document que nous avons implémentées. Celles-ci découlent de certaines analyses plus théoriques et les résultats de ces analyses permettent au-delà de la perspective d'automatisation de produire des connaissances sur le cycle de vie des Guides de Bonnes Pratiques. Cette articulation prend cependant place dans un environnement informatique unique. Nous proposons donc de fournir un environnement logiciel dédié aux Guides de Bonnes Pratiques permettant du point de vue des tâches théoriques, d'analyser et de caractériser des propriétés des Guides de Bonnes Pratiques ; puis d'un point de vue des tâches pratiques, de proposer des outils d'aide à l'encodage, d'aide à la structuration, d'aide à la consultation, d'aide à l'écriture, d'aide à l'extraction de connaissances.

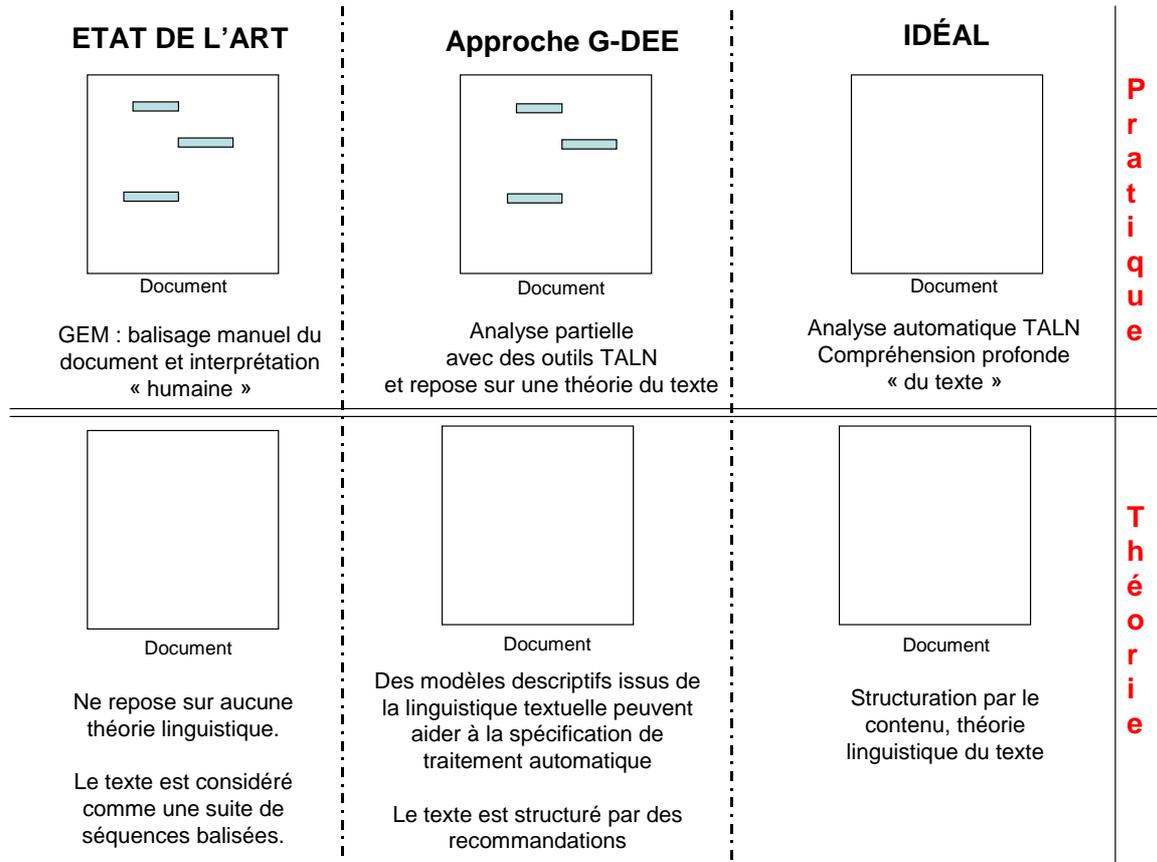


Figure 1.1 – Schéma représentant la stratégie d'ensemble de l'approche adoptée.

La Figure 1.1 présente la stratégie d'ensemble de notre approche articulant les aspects pratiques et théoriques que nous avons adoptés issus des travaux existants en Informatique Médicale et des théories linguistiques que nous décrivons dans le chapitre 3.

1.2 Présentation du document

Après avoir présenté les objectifs de cette thèse et sa philosophie générale, nous décrivons dans cette section l'organisation des différents chapitres.

Le chapitre deux constitue une introduction aux Guides de Bonnes Pratiques, tant au niveau de leur élaboration, que de leur structuration et de leur impact auprès des praticiens. Il peut plus particulièrement intéresser les personnes qui s'attachent à l'élaboration et l'informatisation des Guides de Bonnes Pratiques. Nous présentons dans ce chapitre, les principales approches proposées pour l'informatisation des Guides de Bonnes Pratiques, en les classant en deux catégories principales, l'une centrée sur la formalisation des connaissances, et l'autre centrée sur le traitement du document. Ce chapitre ne prétend pas à l'exhaustivité et nous avons fait le choix de présenter des approches représentatives de chaque type de formalisme de représentation des connaissances, s'intéressant au cycle de vie du document. Nous concluons ce chapitre par une mise en contexte générale de notre approche qui consiste en une extension d'une des principales approches documentaires par la reconnaissance automatique de recommandations contenues dans les Guides de Bonnes Pratiques.

Le chapitre trois est consacré à la caractérisation des phénomènes linguistiques permettant de structurer les Guides de Bonnes Pratiques, et plus particulièrement permettant de reconnaître automatiquement les recommandations contenues dans les Guides de Bonnes Pratiques. Nous introduisons la notion d'opérateurs déontiques dont la forme linguistique (verbes déontiques de type recommander, devoir, envisager) constitue l'expression des recommandations. Pour ce faire, l'étude des champs lexicaux des verbes déontiques ainsi que des analyses distributionnelles des opérateurs déontiques dans les textes seront présentées.

Le chapitre quatre est dédié à la description de notre approche technique, consistant à inclure des techniques empruntées à l'Extraction d'Informations à une approche documentaire plus classique comme principe servant à la structuration automatique.

Le chapitre cinq décrit l'environnement G-DEE (*Guideline Document Engineering Environment*) que nous avons développé. Il s'attache à l'interface graphique, l'utilisation et aux fonctionnalités de l'approche présentée dans cette thèse. G-DEE est dédié à la fois à l'étude des Guides de Bonnes Pratiques et à terme à leur analyse automatique, et intègre de nombreuses fonctionnalités que nous décrivons à travers ses modes d'utilisation.

1.2. Présentation du document

Le chapitre six est dédié aux techniques de traitement linguistique, plus spécifiquement à la description des automates à états finis qui permettent de reconnaître de manière automatique les marqueurs textuels décrits dans le chapitre 3. Ce chapitre intéressera les lecteurs attachés aux aspects informatiques les plus techniques de G-DEE (c'est-à-dire les formalismes d'analyse). Dans une première étape, nous décrivons les pré-traitements nécessaires à l'analyseur, puis nous décrivons les automates à états finis permettant la reconnaissance des marqueurs textuels.

Le chapitre sept décrit le fonctionnement de l'analyseur ainsi que l'algorithme dédié à l'extraction et à l'affichage du texte analysé, puis balisé. Ce chapitre fait le lien entre l'analyse linguistique et les aspects documentaires, représentés par le balisage des textes et différents traitements dérivés de ce balisage, qu'ils soient visuels comme les modalités d'affichage et de mise en valeur de segments textuels, ou conceptuels, comme la dérivation automatique de connaissances formalisées. Nous présentons donc les techniques utilisées basées essentiellement sur XML et les feuilles de style associées XSL¹ (*eXtensible Stylesheet Language*) ainsi que les processeurs XSLT² (*XSL Transformations*) permettant d'interpréter ces données. Les différents traitements présentés dans ce chapitre permettent de surligner le texte correspondant à des recommandations, d'identifier certaines informations du texte pour dériver de manière automatique des éléments GEM, ou des règles de décision.

Le chapitre huit présente certains phénomènes syntaxiques « réels » rencontrés dans les Guides de Bonnes Pratiques et qui sont traités par G-DEE, et est donc l'occasion de confronter le système aux limites de sa performance. Ces différents exemples de phrases représentant des phénomènes linguistiques spécifiques permettront aussi, au-delà de certaines limitations du système, de mettre en évidence certains problèmes de qualité d'écriture des Guides de Bonnes Pratiques qui se manifestent par des structures atypiques.

Le chapitre neuf présente deux expériences d'évaluation de G-DEE. Ce chapitre intéressera les personnes s'attachant aux performances techniques de G-DEE en tant que base d'un outil opérationnel dans le futur, ainsi que les impacts observés auprès d'experts médicaux. La première expérience que nous avons menée consiste à évaluer les performances techniques de G-DEE définies par le taux de reconnaissance des recommandations en comparaison à un balisage manuel d'un ensemble de Guides de Bonnes Pratiques. La deuxième expérience consiste à faire évaluer les Guides de Bonnes Pratiques par des experts intervenant tant au niveau de l'élaboration que de l'évaluation des Guides de Bonnes Pratiques, et impliqués dans la mise en œuvre de ces Guides. Une grille de cotations, inspirée par les évaluations publiées pour les systèmes d'Extraction d'Informations ou de Recherche d'Informations, a été définie. Les résultats seront analysés en ce qui concerne la validité de l'approche et permettront de poser différentes problématiques quant à l'écriture et l'élaboration de ces documents textuels, qui peuvent constituer autant de pistes méthodologiques pour l'amélioration du cycle de vie des Guides de Bonnes Pratiques.

¹ <http://www.w3.org/Style/XSL/> (dernier accès le 10 juin 2006)

² <http://www.w3.org/TR/xslt> (dernier accès le 10 juin 2006)

Enfin, le chapitre dix est dédié aux perspectives d'utilisation de G-DEE après certaines améliorations et évolutions techniques. Ces perspectives sont présentées à travers le prisme de différents « modes d'utilisation » regroupant les fonctionnalités élémentaires en tâches de plus haut niveau intéressant également des populations d'utilisateurs spécifiques. Nous décrivons un ensemble de perspectives possibles à partir de G-DEE et proposant donc un G-DEE étendu pour intervenir à toutes les étapes du cycle de vie des Guides de Bonnes Pratiques, par exemple pour améliorer la rédaction des documents, ou faciliter leur informatisation.

Une annexe présente les publications réalisées durant cette thèse, ainsi que les sources de référence en ce qui concerne les Guides de Bonnes Pratiques, ainsi que des exemples de résultats plus étendus. Nous présentons les Guides de Bonnes Pratiques utilisés pour l'analyse linguistique, et ceux pour constituer la grammaire des automates à états finis, puis nous présentons un Guide de Bonnes Pratiques balisé par G-DEE et évalué par des experts lors de l'évaluation. Enfin, nous présentons la synthèse des commentaires oraux des experts sur certaines portions de texte du Guide de Bonnes Pratiques.

Chapitre 2

Les Guides de Bonnes Pratiques, de leur élaboration à leur informatisation

Ce chapitre constitue une introduction aux Guides de Bonnes Pratiques, tant au niveau de leur élaboration, que de leur structuration et de leur impact auprès des praticiens. Il peut plus particulièrement intéresser les personnes qui s'attachent à l'élaboration et l'informatisation des Guides de Bonnes Pratiques. Nous présentons dans ce chapitre, les principales approches proposées pour l'informatisation des Guides de Bonnes Pratiques, en les classant en deux catégories principales, l'une centrée sur la formalisation des connaissances, et l'autre centrée sur le traitement du document. Ce chapitre ne prétend pas à l'exhaustivité et nous avons fait le choix de présenter des approches représentatives de chaque type de formalisme de représentation des connaissances, s'intéressant au cycle de vie du document. Nous concluons ce chapitre par une mise en contexte générale de notre approche qui consiste en une extension d'une des principales approches documentaires par la reconnaissance automatique de recommandations contenues dans les Guides de Bonnes Pratiques.

2.1 Cycle de vie des Guides de Bonnes Pratiques

Les Guides de Bonnes Pratiques aussi appelés Recommandations pour la Pratique Clinique³ ou Recommandations Professionnelles³ sont des documents textuels médicaux contenant des recommandations élémentaires⁴ (plus communément appelées recommandations)⁵. Celles-ci sont basées sur le concept d'*Evidence-Based Medicine*⁶ ou médecine factuelle (Sackett et al., 1996). Sackett définit l'*Evidence-Based Practice* comme « *the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patient* ». Les Guides de Bonnes Pratiques sont donc définis comme « un ensemble d'instructions spécifiques, méthodologiquement développées afin d'assister le praticien et le patient dans la décision d'un soin approprié, selon des circonstances cliniques spécifiques »⁷ (Field & Lohr, 1990). Cette définition est reprise par le Professeur Y. Matillon, Directeur général de l'Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES)⁸, dans l'avant-propos de toutes les références médicales publiées par l'Agence : « L'objectif principal [des recommandations] est de fournir aux praticiens une synthèse du niveau de preuve scientifique des données actuelles de la science et de l'opinion d'experts sur un thème de pratique clinique, et d'être ainsi une aide à la décision en définissant ce qui est approprié, ce qui ne l'est pas ou ne l'est plus, et ce qui reste incertain ou controversé. ».

Les Guides de Bonnes Pratiques sont généralement élaborés par des sociétés savantes ou des agences nationales, comme l'HAS (Haute Autorité de Santé, anciennement ANAES) pour la France, NICE⁹ (*National Institute for health and Clinical Excellence*) pour l'Angleterre, JNC¹⁰ (*Joint National Committee*) et NGC¹¹ (*National Guideline Clearinghouse*) pour les États-Unis (Keffer, 2001). Ces documents sont distribués en version papier et en version électronique accessibles par Internet (Figure 2.2), et sont généralement actualisés régulièrement (ANAES, 2000; Armstrong et al., 1992) en fonction des progrès scientifiques et ne s'inscrivent donc pas dans un cadre rigide mais définis dans une politique d'adaptation de la médecine par rapport aux avancées de la science (Frankel et al., 1999). Les Guides de Bonnes Pratiques peuvent également être intégrés dans des systèmes d'aide à la décision ou

³ Cette dénomination est essentiellement utilisée par l'HAS (Haute Autorité de Santé)⁸ et correspond au document textuel dans son ensemble.

⁴ La recommandation élémentaire désigne chacune des phrases correspondant à une conduite à tenir pour une situation clinique donnée, par exemple « Il est recommandé d'initier un traitement antihypertenseur, dans un premier temps, par une monothérapie à faible dose, en particulier après 80 ans. ». La structure des Recommandations pour la Pratique Clinique ou Guides de Bonnes Pratiques correspond donc à un assemblage de recommandations élémentaires.

⁵ Nous utilisons le terme de Guides de Bonnes Pratiques et de recommandations dans l'ensemble de ce document.

⁶ <http://www.cebm.net/index.asp> (dernier accès le 10 juin 2006).

⁷ *systematically developed statements to assist practitioner and patient decisions about appropriate Healthcare for specific clinical circumstances.*

⁸ <http://www.has-sante.fr/> (dernier accès le 10 juin 2006).

⁹ <http://www.nice.org.uk/> (dernier accès le 10 juin 2006).

¹⁰ <http://www.nhlbi.nih.gov/guidelines/> (dernier accès le 10 juin 2006).

¹¹ <http://www.guideline.gov/> (dernier accès le 10 juin 2006).

des systèmes d'alerte permettant ainsi de proposer les « bonnes pratiques » selon un profil spécifique de patient au moment de la prise de décision (Gillois et al., 2001).

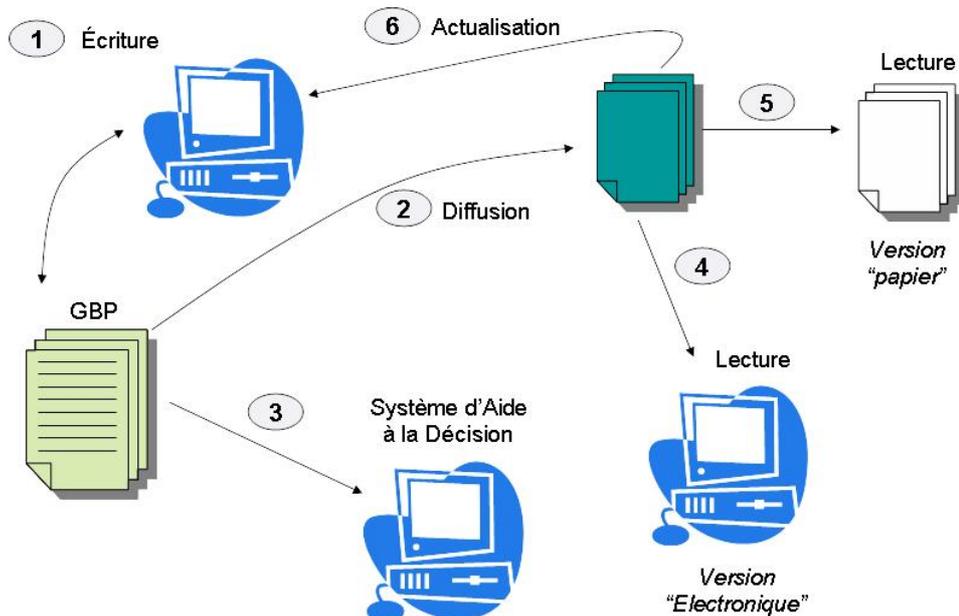


Figure 2.2 – Cycle de vie des Guides de Bonnes Pratiques, de leur élaboration jusqu'à leur diffusion et leur intégration dans des systèmes d'aide à la décision.

Dans les sections suivantes, nous décrivons chaque étape du cycle de vie des Guides de Bonnes Pratiques représentée par la Figure 2.1, tant au niveau de leur élaboration, que de leurs objectifs initiaux, ainsi que les principaux obstacles de leur non-utilisation et de leur non adhérence par les praticiens.

2.2 **Élaboration des Guides de Bonnes Pratiques**

Comme nous l'avons défini précédemment, les Guides de Bonnes Pratiques décrivent les conduites cliniques les plus appropriées pour une pathologie donnée, déterminées en utilisant les données les plus récentes de la littérature médicale, validées par un consensus d'experts spécialement réunis pour l'élaboration d'un Guide de Bonnes Pratiques. Les « revues systématiques » représentent une des méthodes les plus efficaces pour s'assurer que la preuve est évaluée avec le moins d'erreurs possibles ou de biais. Ces revues systématiques peuvent fournir un résumé de la meilleure preuve récente disponible d'un sujet spécifique et sont valables pour clarifier les résultats à travers une large gamme de sources.

Deux approches de rédaction des Guides de Bonnes Pratiques ont été décrites dans la littérature. La méthode *top down* consiste à la confier exclusivement à des experts reconnus, au risque de susciter des critiques de la part des praticiens de « terrain », et à l'inverse la méthode *bottom up* consiste à inclure dans le groupe de travail des médecins « de terrain »,

afin de prendre en compte les opinions des praticiens appelés à appliquer quotidiennement ces recommandations (Armstrong et al., 1992).

Les Guides de Bonnes Pratiques sont des documents généralement pensés comme un outil éducatif pour les professionnels de Santé, et comme une cible pour la formation permanente. Par exemple, comme le décrit Dawes (Dawes, 1996), les « médecins de famille » n'ont généralement pas les compétences et l'expérience nécessaires pour juger les résultats et déterminer s'ils sont applicables à leur exercice. Dans ce contexte, les Guides de Bonnes Pratiques pourraient trouver leur place et aider le praticien à mettre en œuvre un traitement efficace pour leur patient basé sur les dernières informations des essais cliniques et des procédures médicales (Patel et al., 2004). Dans cette étude, Patel et ses collègues ont montré que la majorité des praticiens non spécialistes de la dépression avaient des difficultés pour reconnaître les symptômes de la maladie sans l'utilisation d'un Guide de Bonnes Pratiques. L'intérêt des Guides de Bonnes Pratiques est donc bien avant tout un outil d'aide à la conduite à tenir face à une situation clinique. De la même manière, Farquhar et al. (Farquhar, Kofa & Slutsky, 2003) ont montré que 75% des praticiens considèrent les Guides de Bonnes Pratiques comme des outils éducatifs fournissant de « bons conseils ». Parmi ces praticiens, 70% considèrent que les Guides de Bonnes Pratiques ont pour vocation d'améliorer la qualité des soins, et 53% les décrivent comme un outil de réduction des coûts.

Certains praticiens s'intéressent à l'utilisabilité des Guides de Bonnes Pratiques, tandis que d'autres les considèrent comme un outil pouvant réduire l'autonomie des praticiens et même augmenter les risques d'actions en justice (41%). Néanmoins, l'*Evidence-Based Medicine* n'est pas une « médecine de recettes » (*cook book medicine*) (Sackett et al., 1996), mais devrait intégrer l'expertise clinique individuelle avec la meilleure preuve clinique (*evidence*) issue des recherches systématiques (Hewitt-Taylor, 2003).

2.3 Rôle des Guides de Bonnes Pratiques

On considère aujourd'hui que la grande variabilité des pratiques médicales observée dans plusieurs études (Patel et al., 2001a) peut être préjudiciable à la qualité des soins apportés aux patients (Keckley, 2004). Ainsi, un des principaux objectifs des Guides de Bonnes Pratiques consiste à promouvoir les « bonnes pratiques » pour réduire les variations « inappropriées », et améliorer ainsi la qualité des soins (Farquhar, Kofa & Slutsky, 2002; Grimshaw & Russel, 1993). Les contrastes observés entre ce que devrait être la pratique clinique (conforme à des référentiels) et ce qu'elle est réellement (Bero et al., 1998; Lomas, 1991) montre que le patient ne bénéficie pas toujours de soins validés par les données scientifiques les plus récentes, et qu'il peut donc encourir des risques du fait de pratiques obsolètes ou injustifiées.

Les Guides de Bonnes Pratiques permettent aussi d'agir sur le contrôle de la qualité, en accentuant les imperfections de la littérature existante et pouvant suggérer les « futurs » axes de recherche. La connaissance médicale est sujette à une continuelle croissance et à de nombreuses révisions (Sackett et al., 1996). Un des problèmes décrit par les praticiens

2.3. Rôle des Guides de Bonnes Pratiques

concerne l'actualisation de leurs connaissances parmi les différentes sources de connaissances existantes basées sur les nouvelles preuves scientifiques (McGlynn et al., 2003) du fait de l'avènement des nouvelles technologies de l'information et de la communication. Ainsi, on a pu suggérer que les Guides de Bonnes Pratiques proposent une synthèse de ces évolutions permettant au praticien d'actualiser ses connaissances sans être confronté à la surcharge d'informations découlant de la disponibilité en ligne de nombreuses références : *“Clinical guidelines are meant to help manage this ‘information overload’ and provide up-to-date, scientifically valid information to aid the practitioner in making more efficient and effective clinical decisions”* (Grimshaw & Russel, 1993).

De plus, le coût toujours plus important des soins rend nécessaire le développement de « bonnes pratiques » (Midwest Business Group on Health, 2003), idéalement dans le but de réduire les coûts en augmentant simultanément la qualité des soins ! Une étude de la *Juran Institute* a montré que les coûts de Santé pouvaient être réduits de 30% tout en améliorant la qualité des soins apportés aux patients. Ainsi, l'augmentation des coûts proviendrait pour une large part de soins inappropriés (retards, pathologies iatrogènes, etc.) auquel cas les Guides de Bonnes Pratiques pourraient être un outil pour les réduire. Ce nouvel objectif peut se rapprocher de celui émis vers la fin des années 1970 aux États-Unis avec les Recommandations Médicales dans le cadre du programme « *Consensus Development Conferences* », géré par « *The Office of Medical Application of Research* » (OMAR). En effet, ce sont des préoccupations légales et financières qui ont été à la base des Recommandations Médicales. En France, les Références Médicales Opposables (RMO) diffusées dans les années 80 procédaient d'une logique similaire avant d'être abandonnées. Certaines conduites de références ont ainsi pu devenir des arguments éventuellement opposables aux médecins dont la pratique pouvait s'écarter d'une médecine fondée sur l'évaluation rigoureuse des rapports bénéfice/risque et coût/efficacité. Une autre étude menée par Perry et Marx (Perry & Marx, 1994) a montré que sur 8 organisations élaborant des Guides de Bonnes Pratiques aux États-Unis, toutes évoquaient qu'elles les développaient avec l'objectif d'améliorer l'efficacité clinique et d'améliorer la qualité des soins, sans mettre en avant dans un premier temps aucune justification de type économique. Néanmoins, six d'entre elles ont admis que le but second de ces Guides de Bonnes Pratiques restait le contrôle des coûts. Ainsi, l'impact attendu de ces Guides de Bonnes Pratiques consiste à réduire les pratiques inappropriées et/ou inefficaces en encourageant les pratiques efficaces, améliorant la qualité, et réduisant les coûts de Santé¹² (Grimshaw & Russel, 1993; Lomas, 1991; McAlister et al., 1999).

Les Guides de Bonnes Pratiques jouent également un rôle important dans la diffusion et la standardisation des connaissances médicales elles-mêmes, étant basés sur le concept d'*Evidence-Based Medicine*. Initialement, faiblement diffusés car ils étaient sur support papier, les Guides de Bonnes Pratiques ont beaucoup profité de l'avènement des nouvelles technologies de l'information et de la communication. Ainsi, les versions numériques de ces documents sont actuellement largement accessibles sur les réseaux (Internet, intranet, etc.). Néanmoins, d'autres études ont montré que la dissémination de ces Guides de Bonnes Pratiques à travers l'intranet n'apportait pas de bénéfices par rapport à la version

¹² *reduce inappropriate and/or ineffective practices while encouraging effective practices, improving quality, and reducing costs of health care.*

papier (Stolte, Ash & Chin, 1999). De la même manière, Jousimaa et ses collègues (Jousimaa et al., 2003) ont montré que les Guides de Bonnes Pratiques sont peu utilisés et n'ont pas d'impact significatif lors de la prise de décision par les médecins. Les praticiens récemment qualifiés, ont une meilleure connaissance des Guides de Bonnes Pratiques, et n'hésitent pas à les appliquer en pratique courante. Mais, l'informatisation des Guides de Bonnes Pratiques seule ne semble pas susceptible d'apporter une plus-value à la pratique quotidienne.

La diffusion et l'accessibilité des Guides de Bonnes Pratiques sont des facteurs pour améliorer l'observance des « bonnes pratiques » (Costanza et al., 1992). Néanmoins, la diffusion passive¹³ est reconnue comme étant finalement nuisible à leur bonne observance (Bero et al., 1998; Burstin et al., 1999). La meilleure méthode de diffusion des Guides de Bonnes Pratiques semble être la formation interactive consistant en « une véritable éducation des praticiens, comprenant des séminaires de sensibilisation et d'information » (ANAES, 2000), dans la mesure où l'enseignement dans le cadre de la formation médicale continue ne semble pas suffisant pour améliorer l'observance des praticiens. En ce qui concerne les interventions de « leaders d'opinion » ou d'approches « qualité », les études d'évaluation de leur efficacité sont encore trop rares et discordantes pour aboutir à des conclusions nettes en leur faveur (ANAES, 2000). Il n'en demeure pas moins que la méthode la plus employée reste la publication dans des revues médicales (Lomas, 1991). Bien qu'elle s'apparente encore à la diffusion passive, l'impact des publications joue un rôle important dans l'acceptation des Guides de Bonnes Pratiques par les médecins. Il apparaît que l'utilisation de plusieurs actions de diffusion potentialise les effets de chacune de ces actions. « C'est donc une véritable stratégie de diffusion qui doit être établie par les auteurs ou les commanditaires des Guides de Bonnes Pratiques, afin de leur assurer un impact appréciable » (Caniard, 2002).

2.4 Obstacles à l'utilisation des Guides de Bonnes Pratiques

La simple diffusion des Guides de Bonnes Pratiques n'a qu'un faible impact sur le comportement des médecins (Grimshaw & Russel, 1993; Matillon & Durieux, 2000), et seul l'encadrement des pratiques incluant une recommandation centrée-patient semble influencer la prise de décision du médecin dans le sens d'une augmentation de l'observance des recommandations. De nombreuses études rapportent les raisons de la non-utilisation des Guides de Bonnes Pratiques, qui sont directement corrélées à la connaissance des praticiens, à leur attitude ou à leur comportement (Cranney et al., 2001).

Cabana et al. (Cabana et al., 1999) décrivent les différents obstacles à l'observance des Guides de Bonnes Pratiques parmi toutes les études publiées en s'intéressant plus particulièrement à ceux qui pourraient être la cible d'actions spécifiques :

- méconnaissance (*lack of awareness*),

¹³ La dissémination passive de l'information consiste, par exemple, en des publications de conférences de consensus dans des journaux professionnels ou du courrier à but éducatif.

2.4. Obstacles d'utilisation des Guides de Bonnes Pratiques

- manque de familiarité (*lack of familiarity*),
- désaccord (*lack of agreement*),
- manque de confiance dans la capacité à suivre les Guides de Bonnes Pratiques (*lack of self-efficacy*),
- manque de résultats (*lack of outcome expectancy*),
- force des habitudes (*inertia of previous practice*),
- facteurs externes (*external barriers*), par exemple : les exigences du patient incompatibles avec les recommandations.

Les facteurs susceptibles d'être ciblés sont catégorisés selon l'expertise des praticiens, l'attitude face à la notion de Guides de Bonnes Pratiques, et le comportement en général des praticiens (Cabana et al., 1999).

Les Guides de Bonnes Pratiques se focalisent généralement sur le traitement de pathologies spécifiques, telles que le dépistage du diabète ou de la prise en charge de l'hypertension artérielle, et sont généralement présentés sous forme textuelle décrivant des procédures validées par des essais cliniques. Les Guides de Bonnes Pratiques contiennent généralement des stratégies thérapeutiques (Figure 2.3) sous forme de représentations algorithmiques organisant toutes les informations pertinentes dans un format directement applicable, et guidant ainsi la prise de décision (Patel et al., 2001b).

Ces documents sont souvent trop longs pour pouvoir être utilisés en pratique et sont parfois complexes et peu clairs (Patel et al., 2001a). Des études ont montré le manque d'une structure standardisée des Guides de Bonnes Pratiques publiés. Ils sont produits et disséminés par de nombreux organismes publics et suivent des structures assez variables ce qui peut rendre leur compréhension difficile (Elkin et al., 2000).

C'est pourquoi, de nombreuses études se sont focalisées sur l'écriture et la standardisation des Guides de Bonnes Pratiques, en ayant notamment recours à leur informatisation pour soutenir cet effort de standardisation. Récemment, Shiffman et ses collègues (Shiffman et al., 2003) concluaient qu'un des principaux facteurs qui affectait la qualité des Guides de Bonnes Pratiques provenait du fait que le développement des Guides de Bonnes Pratiques souffrait du manque de formation spécifique de leurs auteurs à la rédaction de documents normatifs. Ces textes peuvent ainsi être complexes sémantiquement, et sur le plan structurel, contiennent souvent des séquences de procédures comprenant des relations causales ou temporelles également complexes (Patel et al., 2001a).

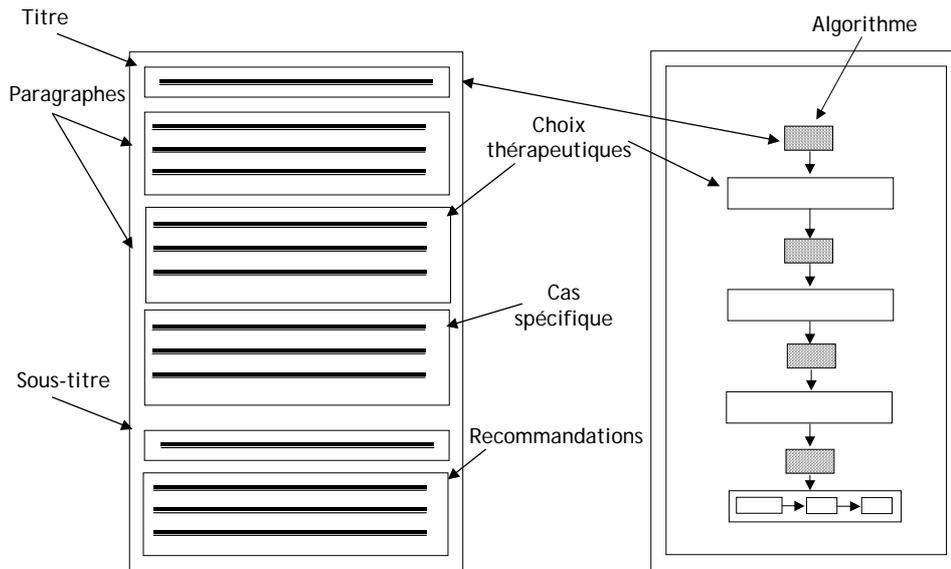


Figure 2.3 – Structure du Guide de Bonnes Pratiques composé d’une partie textuelle et illustré par des algorithmes représentant généralement la stratégie thérapeutique.

Outre la complexité, l’ambiguïté est un autre obstacle à la compréhension des Guides de Bonnes Pratiques. Ces problèmes d’ambiguïtés sont décrits dans de nombreuses études (Basili, Grisoli & Pazienza, 1994; Codish & Shiffman, 2005) ; dans le cas des Guides de Bonnes Pratiques on peut faire l’hypothèse d’écriture intentionnellement ambiguë dans le but d’arriver au consensus indispensable à leur acceptation par le groupe de travail les rédigeant (Tierney, 1995). Une meilleure compréhension de la sémantique des Guides de Bonnes Pratiques pourrait améliorer leur qualité, et faciliter leur utilisation (Hewitt-Taylor, 2004; Patel et al., 2001a; Shiffman & Michel, 2004; Shiffman et al., 2003).

Les Guides de Bonnes Pratiques peuvent être caractérisés par leur structure et la connaissance spécifique qui y est contenue. Leur forme logique a inspiré de nombreuses recherches dans le domaine du traitement du document et plus particulièrement dénommées « approche centrée sur le document » (*document-centered approach*), alors que la connaissance contenue dans les Guides de Bonnes Pratiques sert de point de départ à des systèmes d’aide à la décision, appelé aussi « approche centrée sur les connaissances » (*knowledge-based approach*). Les recherches portant sur leur structure constitue donc une pré-condition pour leur informatisation (Shiffman & Michel, 2004).

Nous présentons ci-dessous une synthèse des principaux problèmes soulevés par les Guides de Bonnes Pratiques.

2.4. Obstacles d'utilisation des Guides de Bonnes Pratiques

Tableau 2.1 – Tableau représentant les problèmes issus de leur forme textuelle.

Problème identifié	Approches possibles pour le résoudre
Complexité sémantique	Contrôler la rédaction, la normalisation.
Structure variable	Standardiser la structure et l'élaboration.
Difficulté à les mettre en pratique	Informatiser les Guides de Bonnes Pratiques dans des systèmes d'aide à la décision, des systèmes d'aide à la lecture.
Difficulté d'élaboration	Fournir une aide à l'écriture par la détection des recommandations et de leur représentation sous une forme standardisée du type « si ... alors ».

Tableau 2.2 – Tableau représentant les problèmes issus de leur élaboration.

Problème identifié	Approches possibles pour le résoudre
Cible des recommandations (patients/médecins)	Élaborer les Guides de Bonnes Pratiques en fonction de leur cible.
Synthèse des études de référence (par exemple essais cliniques) dans le Guide de Bonnes Pratiques	Développer des méthodes de synthèse.
Manque de clarté et ambiguïté dérivés du consensus	Développer des méthodes formelles de consensus.

Tableau 2.3 – Tableau représentant les problèmes de non-utilisation des Guides de Bonnes Pratiques par les praticiens.

Problème identifié	Approches possibles pour les résoudre
Les Guides de Bonnes Pratiques sont trop prescriptifs (« recettes »)	Justifier et définir les conditions d'applicabilité de manière plus apparente dans le document.
Longueur et complexité des documents	Mieux structurer le document ; décomposition standard des conduites à tenir.
Inertie des pratiques	Démontrer les avantages des recommandations en pratique, et du bénéfice/risque des pratiques usuelles.

2.5 Informatisation des Guides de Bonnes Pratiques : approches centrées sur les connaissances

La diffusion des Guides de Bonnes Pratiques constitue une étape importante de leur cycle de vie (Figure 2.2). L'informatisation des Guides de Bonnes Pratiques représente une innovation importante par rapport à leur diffusion sur support papier ; au-delà de la démultiplication de leur diffusion elle peut faciliter différents types de traitements susceptibles d'améliorer leur qualité. On peut citer en particulier :

- (i) la détection des erreurs contenues dans les Guides de Bonnes Pratiques ;
- (ii) l'amélioration de la clarté du document, par exemple l'identification de critères de décision et de recommandations cliniques ;
- (iii) la génération automatique en temps réel d'aide à la décision ou d'alertes.

Parallèlement, l'intégration du contenu des Guides de Bonnes Pratiques dans différents systèmes informatiques à finalité clinique est susceptible d'améliorer la qualité de la démarche diagnostique ou thérapeutique.

Par exemple, il a été montré que l'intégration des connaissances des Guides de Bonnes Pratiques dans les feuilles d'admission informatisées permettait de réduire le nombre d'examen demandés dans les unités de soins coronaires sans affecter les résultats attendus pour le patient (Wang et al., 2003). Les outils informatiques intégrant des Guides de Bonnes Pratiques peuvent avoir un réel impact sur le comportement des praticiens pour certaines procédures médicales. D'autres études (Burstin et al., 1999; Caniard, 2002) ont montré que l'intégration des Guides de Bonnes Pratiques à des protocoles issus des établissements de Santé améliorait la prise en charge des patients. De la même manière, l'inclusion des Guides de Bonnes Pratiques dans des systèmes d'aide à la décision améliore l'observance des recommandations par les professionnels de Santé (Garg et al., 2005). Les systèmes d'alertes sont aussi mieux acceptés par les praticiens (Agrawal & Mayo-Smith, 2004; Bindels et al., 2000) et réussissent à modifier le comportement des médecins (Bennett & Glasziou, 2003).

Lobach et Hammond (Lobach & Hammond, 1997) montrent que les systèmes d'aide à la décision basés sur les Guides de Bonnes Pratiques constituent un outil efficace pour assister les professionnels de Santé dans le traitement des patients diabétiques. L'impact du système d'aide à la décision sur le comportement des médecins a été mesuré à travers le taux d'observance des recommandations : celui-ci était de 32% pour le groupe utilisant le système contre 15,6% pour le groupe contrôle.

L'informatisation des Guides de Bonnes Pratiques permet de fournir des outils utilisables en pratique facilitant l'accès aux connaissances, et à ce titre se profile comme une solution au problème de leur mise en œuvre. L'approche « manuelle » consistant à consulter les Guides de Bonnes Pratiques et à se remémorer les différentes recommandations est irréaliste au vu de l'effort qu'elle exige (Grimshaw & Eccles, 2004; Kavanagh, 2002; Patel et al., 2001a).

2.5. Informatisation des Guides de Bonnes Pratiques

Dès les premières tentatives d'informatisation des Guides de Bonnes Pratiques (qui consistaient initialement à créer des systèmes experts), on a pu observer que ces documents contenaient des définitions imprécises et incomplètes (Shiffman & Greenes, 1994). D'après Shiffman (Shiffman et al., 1999), le passage de la forme textuelle d'un Guide de Bonnes Pratiques vers des systèmes informatiques reste complexe. De plus, les Guides de Bonnes Pratiques sont souvent perçus comme des textes n'incluant pas toutes les situations possibles explicitement, ou fournissant des actions alternatives pour le même antécédent au moment de leur formalisation. En conséquence, il est essentiel de vérifier la structure interne du Guide de Bonnes Pratiques avant de l'informatiser en utilisant par exemple les tables de décision (Shiffman, 1997) qui rendent explicites les différentes situations cliniques décrites par le Guide de Bonnes Pratiques (les tables de décision représentent en quelque sorte la « sémantique » des Guides de Bonnes Pratiques).

Dans cette section, nous présentons les principales approches¹⁴ dédiées à la formalisation des connaissances médicales contenues dans les Guides de Bonnes Pratiques. Nous décrivons dans la section suivante celles qui se sont plus particulièrement intéressées aux étapes en amont de cette formalisation, i.e. l'encodage (Georg, 2005).

De nombreux auteurs ont proposé des modèles de représentation des connaissances médicales pour décrire le processus décisionnel contenu dans les Guides de Bonnes Pratiques. Afin de pouvoir comparer les différentes approches, nous nous baserons sur un même exemple, portant sur la prise en charge de la toux chronique présentée dans une étude récente pour comparer les formalismes utilisés dans l'informatisation des Guides de Bonnes Pratiques (Peleg et al., 2003). Un extrait en est présenté ci-dessous :

- (i) La toux chronique est une toux qui dure depuis au moins 3 semaines¹⁵.
- (ii) Les radiographies pulmonaires devraient être effectuées avant de prescrire un traitement chez presque tous les patients porteurs d'une toux chronique (Grade II-2)¹⁶.
- (iii) Les radiographies pulmonaires ne doivent pas être effectuées systématiquement avant le début du traitement : en cas de suspicion de PNDS (syndrome d'écoulement post-nasal), chez les jeunes non fumeurs, les femmes enceintes, ou dans les 4 semaines qui suivent l'arrêt d'un traitement par IEC (Inhibiteur de l'Enzyme de Conversion) chez les patients qui ont développé une toux après l'instauration d'un traitement par IEC¹⁷.

¹⁴ L'exhaustivité des approches est consultable sur le site d'*OpenClinical* ; <http://www.openclinical.org> (dernier accès le 10 juin 2006).

¹⁵ *Chronic cough is cough that lasts for at least 3 weeks.*

¹⁶ *Chest radiographs should be ordered before any treatment is prescribed in nearly all patients with chronic cough (Grade II-2).*

¹⁷ *Chest radiographs do not have to be routinely obtained before beginning treatment, for presumed PNDS (post nasal drip syndrome) in young nonsmokers, in pregnant woman, or before observing the result of discontinuation of an ACE-I (ACE Inhibitor) for 4 weeks for patients who developed cough shortly after beginning to take an ACE-I.*

2.5.1 La Syntaxe Arden et les MLM

La Syntaxe Arden¹⁸, un des langages les plus connus pour représenter et standardiser la connaissance médicale, a été développé dans les années 1980 et a été accepté comme standard par l'ASTM¹⁹ en 1992 et plus tard par l'HL7²⁰ (Jenders, Corman & Sasgupta, 2003). La version initiale de la Syntaxe Arden a été à la base de nombreux schémas d'encodages pour des systèmes experts tels que celui utilisé dans le système HELP (*Health Evaluation through Logical Processing*) à l'hôpital LDH dans l'Utah (Gardner, Pryor & Warner, 1999; Kuperman, Gardner & Pryor, 1991). Ce système rendu disponible aux professionnels de Santé au début des années 1980 permettait de signaler aux praticiens des résultats de laboratoire anormaux ou des interactions médicamenteuses potentiellement dangereuses (Elkin et al., 2000).

La Syntaxe Arden est un langage de description de règles de décision qui permet l'écriture de modules indépendants spécialisés (MLM pour *Medical Logic Modules*) ; plus qu'un simple formalisme de représentation, son ambition a toujours été de permettre de partager la connaissance médicale (Pryor & Hripcsak, 1993). Le praticien, avec peu ou pas de compétences informatiques, peut créer des MLM pour produire des alertes tels que indication de résultats de laboratoires anormaux, interactions médicamenteuses, interprétations diagnostiques, etc. (Anand et al., 2004; Jenders & Dasgupta, 2002).

Parce qu'elle se préoccupe de standardisation, la Syntaxe Arden peut être naturellement utilisée pour représenter le contenu des Guides de Bonnes Pratiques. Néanmoins, les MLM de la Syntaxe Arden ne s'appliquent qu'à des unités de connaissances élémentaires et ne sont pas toujours adaptés aux situations complexes décrites par les Guides de Bonnes Pratiques. En effet, chaque MLM est écrit pour se comporter comme une simple règle, pour laquelle des instructions s'exécutent séquentiellement jusqu'à ce qu'un résultat spécifique soit atteint.

Un MLM est une unité structurée qui décrit une prise de décision élémentaire constituée de trois parties : (1) la déclaration et l'acquisition des données, (2) l'inférence des données (sous la forme de règles SI ... ALORS ...), et (3) l'action à réaliser. Ces MLM sont encodés à l'aide d'une interface graphique dédiée (Bang & Eriksson, 1997).

Trois groupes de propriétés (*slots*) composent le MLM : mise-à-jour (*maintenance*), bibliothèque (*library*) et connaissance (*knowledge*) (Figure 2.4). La catégorie mise-à-jour (*maintenance*) contient des informations descriptives telles que le titre, l'auteur, la version, la date de modification, pour faciliter la mise à jour du MLM et sa portabilité dans les systèmes d'information qui les utilisent. La bibliothèque (*library*) donne des indications concernant l'objectif du MLM ainsi que des mots-clés permettant de le rechercher rapidement. La catégorie connaissance (*knowledge*) contient quatre propriétés (*slots*) :

- Données (*data*) : contient les concepts utilisés dans le MLM,
- Évocation (*evoke*) : représente la condition devant être vérifiée pour que le MLM devienne actif,
- Logique (*logic*) : représente la condition d'exécution (partie SI d'une règle),

¹⁸ <http://cslxinfmtcs.csmc.edu/hl7/arden/> (dernier accès le 10 juin 2006).

¹⁹ ASTM : American Society for Testing and Materials ; <http://www.astm.org> (dernier accès le 10 juin 2006).

²⁰ <http://www.hl7.org> (dernier accès le 10 juin 2006).

2.5.1. La Syntaxe Arden

- Action (*action*) (partie ALORS d'une règle) : déclenche soit une alerte (par exemple, pour administrer un examen de prévention pour le suivi) soit fournit une recommandation spécifique extrait d'un Guide de Bonnes Pratiques (par exemple, le dosage d'une médication) quand la propriété logique est vraie. Un événement déclencheur peut être aussi la saisie d'un diagnostic ou une donnée d'un patient (tels que les résultats des tests de laboratoire).

```
maintenance:
  title: Screen for hypokalemia with digoxin therapy;;
  filename: hypokalemia_and_digoxin;;
  version: 1.06;;
  institution: Columbia-Presbyterian Medical Center;;
  author: George Hripcsak, M.D. (hripcsa@cucis.columbia.edu);;
  specialist: George Hripcsak, M.D.;;
  date: 1993-09-17;;
  validation: production;;

library:
  purpose: Warn the health care provider of hypokalemia in the
  setting of digoxin therapy.;;
  explanation: Whenever a serum or whole blood potassium value is
  stored, it is checked for hypokalemia (less than 3.3).;;
  keywords: hypokalemia; digoxin; arrhythmia ;;
  citations:
    1. International Committee of Medical Journal Editors.
    Uniform Requirements for Manuscripts Submitted to Biomedical
    Journals. NEJM 1991;324:424-8.;;
  links: CTIM-1.14.5;;

knowledge:
  type: data-driven;;
  data:
  K := read last
  where it occurred before now;;
  priority: 50;;
  evoke: potassium_storage;;
  logic:
    if K >= 3.3 then
      conclude false;
    endif;;
  action:
    write "This patient has
    hypokalemia in the setting of
    digoxin therapy...." ;;
    urgency: 50;;
  end:
```

Figure 2.4 – Exemple d'un MLM¹⁸ pour représenter l'alerte de l'hypokaliémie ("*Screen for hypokaliemia with digoxin therapy*").

La Syntaxe Arden permet donc de standardiser et de structurer la connaissance médicale. Cette formalisation, pouvant être faite par des experts médicaux, permet essentiellement de générer des alertes, mais n'est pas adaptée au contenu des Guides de Bonnes Pratiques souvent complexe et devant inclure des imbrications successives.

2.5.2 Le Format d'échange de Guides de Bonnes Pratiques (*Guideline Interchange Format* ou GLIF)

Le formalisme GLIF²¹ (Greenes et al., 1999) a été développé en collaboration avec les universités de Columbia, Stanford, et Harvard (travaillant ensemble à l'*Intermed Collaboratory*²²) s'inspirant des formalismes utilisés dans quatre systèmes d'aide à la décision différents (MLM, MBTA, GEODE-CM et EON) (Ohno-Machado et al., 1998 Jul - Aug). GLIF, considéré comme un modèle objet, reprend les éléments communs à ces quatre systèmes, avec là aussi comme objectif de faciliter le partage des connaissances (Boxwala et al., 2001), contenues dans les Guides de Bonnes Pratiques, à travers différentes institutions et différents systèmes informatiques. Ce modèle se comporte comme un méta-modèle pour décrire et représenter les composants des Guides de Bonnes Pratiques.

Extension des diagrammes et description des processus de décision

La première version de GLIF (Ohno-Machado et al., 1998 Jul - Aug) a été publiée en 1998 et permet de modéliser les Guides de Bonnes Pratiques sous la forme de *diagrammes* selon une syntaxe spécifique. Pour faire face à la complexité des Guides de Bonnes Pratiques, le diagramme permet de décrire l'ordonnancement temporel des étapes le composant (*GuidelineStep*). Ces étapes sont décrites par :

- des conditions à évaluer (*Conditional Step*) et équivalentes à la partie « SI » d'une règle de décision,
- des actions (*Action Step*) équivalentes à la partie « ALORS » de la règle,
- de la simultanéité (*Branch Step*) qui permet de spécifier l'exécution des différentes étapes,
- de la synchronisation (*Synchronization Step*) qui permet d'exécuter les différentes branches.

Les limites de l'approche GLIF

Différentes études sur GLIF (Elkin et al., 2000) ont montré que le processus de structuration de l'information des Guides de Bonnes Pratiques en GLIF requiert des compétences équivalentes à des connaissances informatiques (du fait du caractère algorithmique de la représentation) ; en conséquence, les utilisateurs ont besoin d'un temps significatif pour apprendre le langage (Figure 2.5). Sans cela, les personnes en charge de l'informatisation des Guides de Bonnes Pratiques ont des difficultés à créer des diagrammes GLIF, ce qui entraîne des variations substantielles lors de l'encodage des Guides de Bonnes Pratiques dans le format GLIF (Elkin et al., 2000).

Trois niveaux sont proposés pour encoder les Guides de Bonnes Pratiques avec GLIF (Peleg et al., 2000) :

- (1) le niveau conceptuel : représentation sous forme de diagramme (graphe orienté) ;

²¹ http://www.glif.org/glif_main.html (dernier accès le 10 juin 2006).

²² <http://smi.stanford.edu/projects/intermed-web/Home.htm> (dernier accès le 10 juin 2006).

2.5.2. GLIF

- (2) le niveau de spécification (algorithme générique) qui est automatiquement vérifié pour son uniformité ;
- (3) le niveau d'implémentation.

Des études ont également montré que l'intégrité du texte des Guides de Bonnes Pratiques peut souvent être perdue en les encodant avec GLIF (Patel et al., 1998 Sep-Oct). Une liste non-exhaustive d'imperfections a donc été établie (Greenes et al., 2001; Peleg et al., 2000) :

- (i) GLIF ne permet pas de spécifier la structure de certains attributs importants des Guides de Bonnes Pratiques, par exemple les expressions qui correspondent à des conditions logiques, ou à des valeurs d'attributs exprimées sous la forme de chaînes de caractères qui ne permettent pas une exécution automatique des connaissances formalisées ;
- (ii) l'intégration avec des systèmes cliniques reste difficile ;
- (iii) la sémantique de GLIF reste un mélange de simultanéité et de prise de décision ;
- (iv) des concepts importants manquent, par exemple ceux qui décrivent les itérations, l'état du patient, les états d'exceptions, et les événements.

Pour faire face à ces problèmes, une nouvelle version, GLIF3, ajoute des spécifications à GLIF dont un des buts consiste aussi à assurer l'actualisation des versions des Guides de Bonnes Pratiques encodés avec GLIF (Peleg & Kantor, 2003).

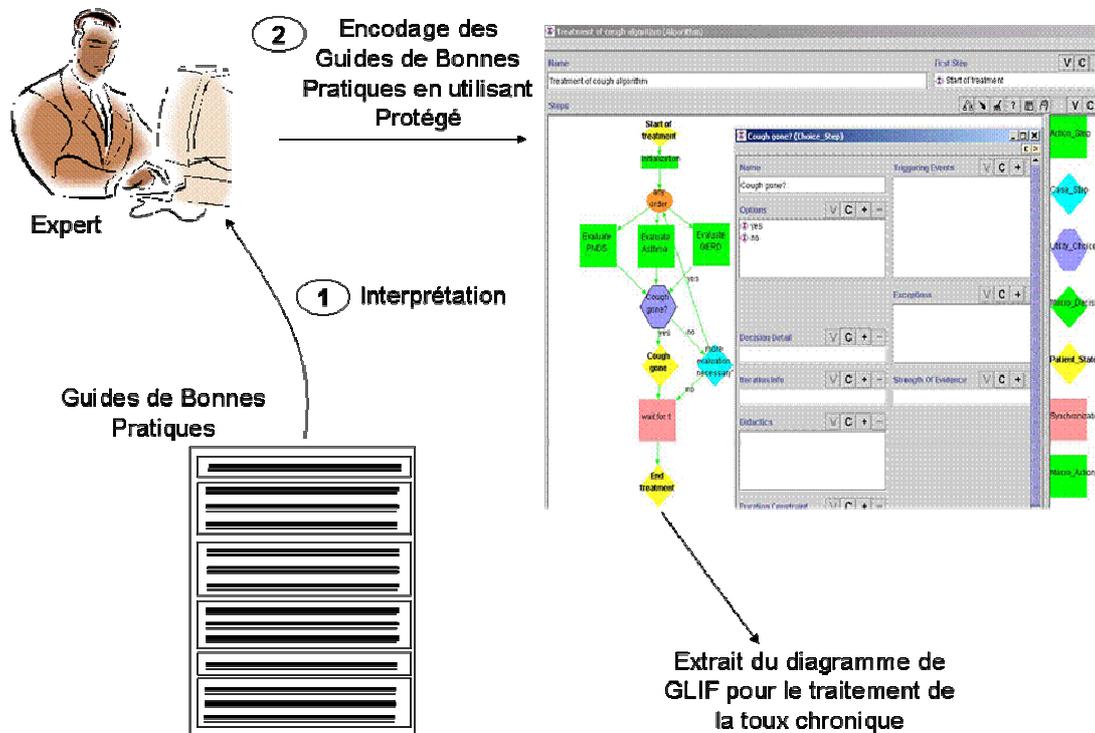


Figure 2.5 – Encodage GLIF de Guides de Bonnes Pratiques en utilisant Protégé²³.

²³ <http://protege.stanford.edu/> (dernier accès le 10 juin 2006).

Le diagramme conceptuel de GLIF3

Les diagrammes initialement décrits dans GLIF ont été améliorés en utilisant les diagrammes de classe de la méthode d'analyse UML²⁴ dans le but d'exprimer la structure et l'organisation des étapes pour supporter des mécanismes imbriqués, essentiellement dans le cas des Guides de Bonnes Pratiques complexes (par exemple, l'imbrication d'autres Guides de Bonnes Pratiques). Ces diagrammes de classe fournissent un modèle de décision flexible à travers une hiérarchie d'étapes de décisions (Figure 2.6). Par exemple, l'*Action step* peut être explorée, du fait de l'imbrication de différents mécanismes permettant de montrer les détails sous la forme d'un autre diagramme. Cette hiérarchie de décision permet de faire une distinction entre les étapes de décision qui peuvent être automatisées (*Case steps*) et celles qui sont spécifiées par le praticien (*Choice steps*).

Extension du diagramme de représentation dans GLIF3

De nouveaux éléments de représentation ont également été développés (cf. Figure 2.7) tels que ceux décrivant (i) les itérations et les conditions qui contrôlent les flux d'itérations ; (ii) les événements et le déclenchement des actions par ces événements ; (iii) les exceptions dans les flux de Guides de Bonnes Pratiques ; (iv) l'état du patient comme dans les étapes du Guide de Bonnes Pratiques (un nœud du diagramme). De cette façon, l'état du patient sert comme point d'entrée dans le Guide de Bonnes Pratiques (de la même manière que dans PRODIGY, décrit plus loin).

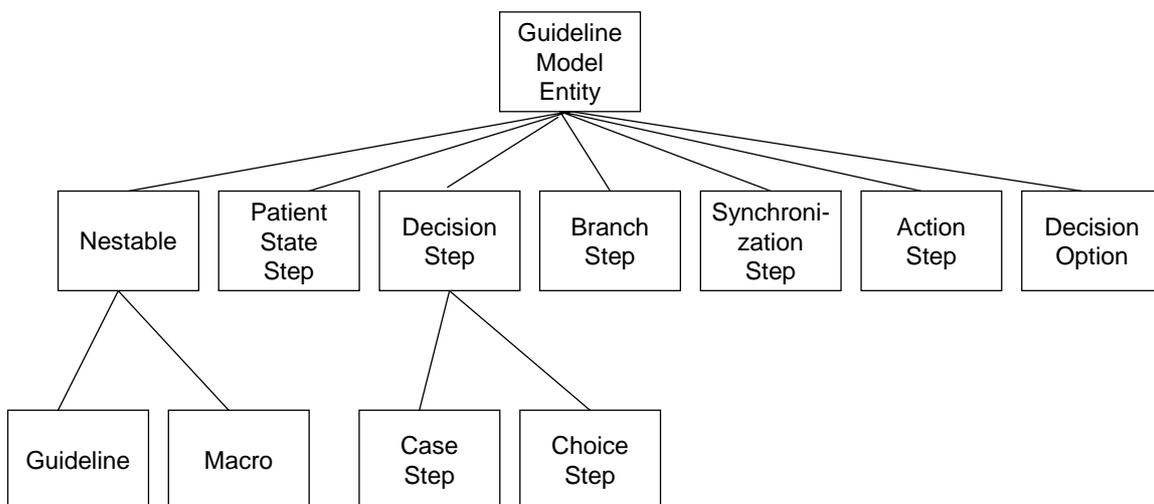


Figure 2.6 – Vue globale des principales classes de la version 3 de GLIF.

Le niveau de spécification

Le niveau de spécification se situe entre le niveau abstrait du diagramme (supporté par la version précédente de GLIF) et le niveau d'implémentation (actuellement partiellement supporté par GLIF3). Le but de ce niveau abstrait du diagramme consiste à aider les auteurs et les utilisateurs à visualiser et à comprendre la structure du Guide de Bonnes Pratiques. Le modèle de spécification a été étendu pour inclure deux types d'actions :

²⁴ *Unified Modeling Language*

2.5.2. GLIF

- (1) les actions pertinentes d'un diagramme (*guideline-flow-relevant actions*) permettant de faire appel à un autre Guide de Bonnes Pratiques, ou de représenter les valeurs de certains paramètres cliniques ;
- (2) les actions pertinentes cliniquement (*clinically relevant actions*) permettant de représenter les actions cliniquement pertinentes qui sont reliées à une ontologie du domaine, telle que celles concernant les prescriptions, les tests de laboratoire, ou les références.

En prenant l'exemple de la toux chronique représentée par la Figure 2.7, on peut observer que le diagramme prend en compte initialement l'état du patient (*Patient State Step*), pour lequel une toux chronique est diagnostiquée. L'étape suivante représentée consiste à questionner le patient (*Action Step*) et vérifier si cette toux n'est pas un effet secondaire de la prise d'IEC (ou ACE sur le diagramme) (*Choice Step*). Si c'est le cas, le traitement est arrêté pendant 4 semaines et le patient est réévalué, sinon on continue l'exploration.

Le diagramme de GLIF a donc l'avantage d'être suffisamment explicite pour être compris et validé par des experts et reflète les différentes conduites à tenir qui sont décrites dans le Guide de Bonnes Pratiques de manière algorithmique. Néanmoins, poussé à l'extrême, ce formalisme pourrait enlever toutes les nuances textuelles contenues dans le Guide de Bonnes Pratiques.

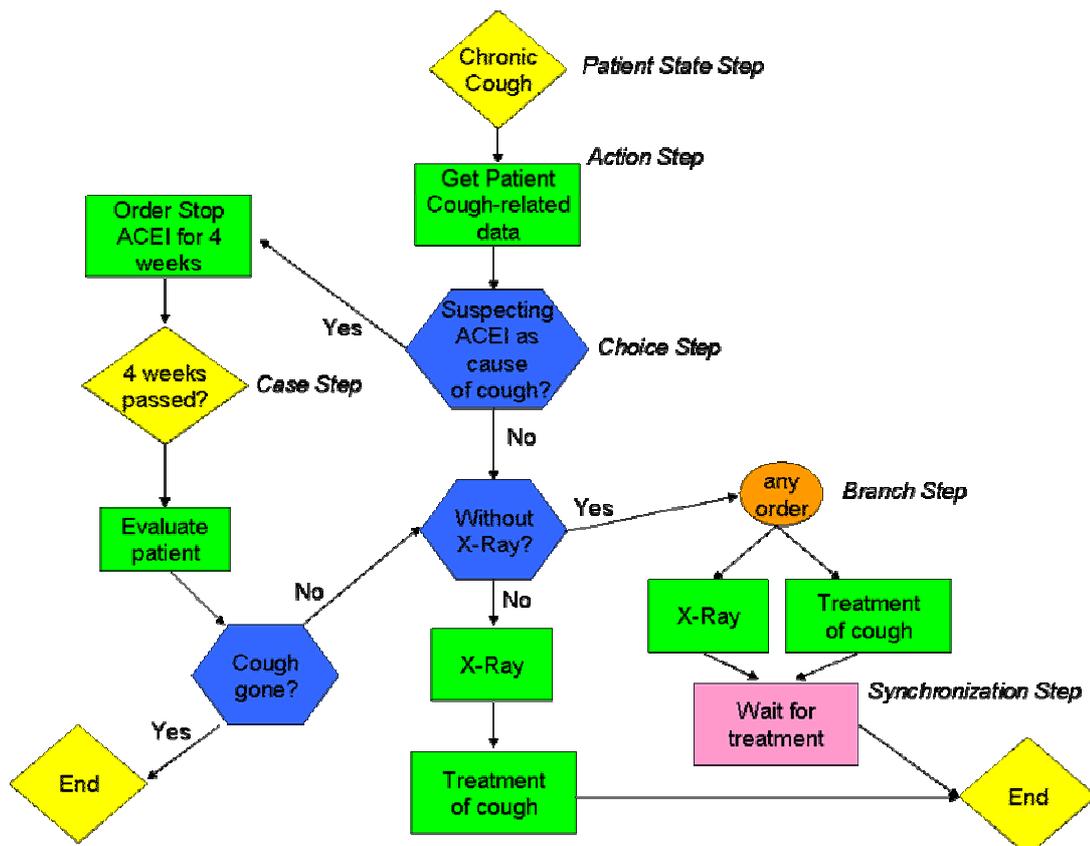


Figure 2.7 – Vue globale des principales classes de la version 3 de GLIF pour l'exemple du traitement de la toux chronique²⁵.

²⁵ <http://smi-web.stanford.edu/auslese/smi-web/reports/SMI-2003-0968.pdf> (dernier accès le 10 juin 2006).

Le niveau d'implémentation

Une syntaxe formelle, basée sur un sur-ensemble de la grammaire de la Syntaxe Arden, a été ajoutée au modèle pour spécifier les expressions et les critères inclus dans les Guides de Bonnes Pratiques (Peleg et al., 2000). Ainsi, de nouveaux opérateurs ont été ajoutés tels que « *is a* », « *overlaps* », « *xor* », « *from now* », « *is unknown* », et « *at least k of* », etc. Un MLM (cf. la section dédié précédemment à la Syntaxe Arden) peut alors être décrit en utilisant un motif des composants de GLIF, principalement pour le relier avec les connaissances issues des Guides de Bonnes Pratiques encodés en GLIF. La nouvelle version de GLIF intègre une ontologie du domaine considérée principalement pour normaliser les termes encodés (Peleg et al., 2001). Elle rejoint en cela d'autres travaux (De Clercq, 2000; Kumar et al., 2004; Pisanelli, Gangemi & Steve, 1999) s'intéressant aux ontologies dans le contexte des Guides de Bonnes Pratiques. Néanmoins, le travail sur l'ontologie du domaine dans GLIF3 est encore en cours de développement. Trois couches constituent actuellement cette ontologie :

- (1) la couche centrale (*Core layer*) correspond à l'interface standard décrivant les données et les concepts médicaux qui peuvent être représentés et référencés par GLIF ;
- (2) la couche du modèle de référence d'information (*Reference Information Model (RIM) Layer*) fournit une hiérarchie sémantique pour les concepts médicaux, et une spécification des attributs pour chaque classe des données médicales ;
- (3) la couche des connaissances médicales (*Medical Knowledge Layer*) contient les termes du dictionnaire (par exemple, la terminologie UMLS²⁶) et peut fournir un accès aux bases de connaissances médicales.

2.5.3 PROforma

PROforma est un langage de représentation des connaissances supporté par des outils d'acquisition et d'exécution des connaissances dont le but consiste à diffuser les Guides de Bonnes Pratiques sous la forme de systèmes experts fiables, fournissant une aide active lors d'une prise en charge médicale, et permettant de gérer les différentes tâches d'un processus de soin (*workflow management*) (Fox & Rahmzadeh, 1998). PROforma a été développé par l'*Advanced Computation Laboratory of Cancer Research*, UK. Le nom de PROforma correspond à une concaténation des termes suivants, `proxy` pour `authorised to act for another` et `formalise` pour `give definite form to`.

Le modèle d'agent autonome « Domino » adapté aux Guides de Bonnes Pratiques

PROforma est basé sur le modèle d'agents autonomes « Domino » (Figure 2.8) et inclut différents composants tels que : les buts, les situations (données patients), les actions, les solutions candidates, les décisions (diagnostic, traitement), et les plans (traitement et plans de soin). Ensemble, ces composants permettent de modéliser le processus des Guides de Bonnes Pratiques (Fox, Beveridge & Glasspool, 2003).

²⁶ <http://www.nlm.nih.gov/research/umls/> (dernier accès le 10 juin 2006).

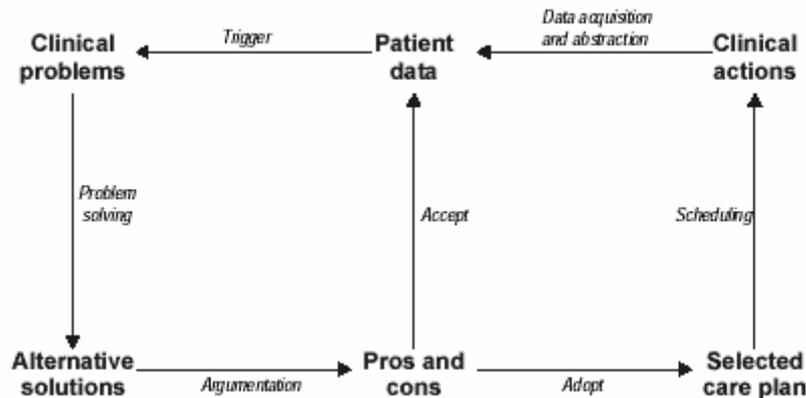


Figure 2.8 – Représentation du modèle Domino dans PROforma.

Ce modèle s’inspire du calcul des prédicats augmenté par des logiques non standards, incluant une syntaxe formelle (Sutton & Fox, 2003 Sep-Oct), définie selon une forme BNF (*forme de Backus-Naur*) pour lesquelles des applications peuvent être développées en utilisant des éditeurs de texte et des compilateurs. PROforma combine la programmation logique et les modèles orientés objets et intègre le langage R²L qui étend la syntaxe standard et la sémantique de Prolog pour utiliser des agents, incluant des décisions, des plans, des actions, des raisonnements temporels et des contraintes. Là où GLIF présentait une vision graphique et algorithmique, PROforma s’apparente davantage à un langage de programmation de haut niveau.

Encodage des Guides de Bonnes Pratiques avec PROforma

De la même manière que pour le modèle GLIF, PROforma représente les Guides de Bonnes Pratiques selon un graphe orienté dans lequel les nœuds sont des instances d’un ensemble de classes, appelé ontologie de tâches PROforma. Chaque Guide de Bonnes Pratiques représenté dans PROforma est modélisé comme un plan constitué d’une séquence de tâches.

Le modèle de tâches PROforma (Figure 2.9) est subdivisé à partir de la tâche générique (*Keystone*) selon quatre types : action (*Action*), plan (*Plan*), décision (*Decision*), et requête (*Enquiry*). Le type *action* correspond à une procédure devant être effectuée, et pouvant inclure la présentation d’informations ou d’instructions que l’utilisateur doit réaliser, par exemple l’envoi d’*e-mail* de rappel, etc. Le type *plan* contient le bloc de base pour l’implémentation des Guides de Bonnes Pratiques et peut contenir soit plusieurs tâches, soit d’autres plans. Le type *décision* correspond à une tâche représentant les choix qui doivent intervenir dans un processus clinique. Le type *requête* quant à lui correspond à une demande d’informations ou de données complémentaires requises pour l’exécution d’un Guide de Bonnes Pratiques. Toutes les tâches partagent des propriétés qui permettent de contrôler leur exécution :

- (i) le but (*goal*) représente les conditions logiques qui une fois vérifiées entraînent l’arrêt de la tâche ;
- (ii) les flux de contrôle (*trigger*) correspondent aux messages qui forcent le déclenchement de la tâche ;

- (iii) les pré-conditions (*preconditions*) représentent les conditions logiques booléennes qui doivent être vérifiées afin que la tâche soit exécutée ;
- (iv) les post-conditions (*postconditions*) représentent les conditions logiques booléennes qui doivent être vérifiées une fois la tâche achevée.

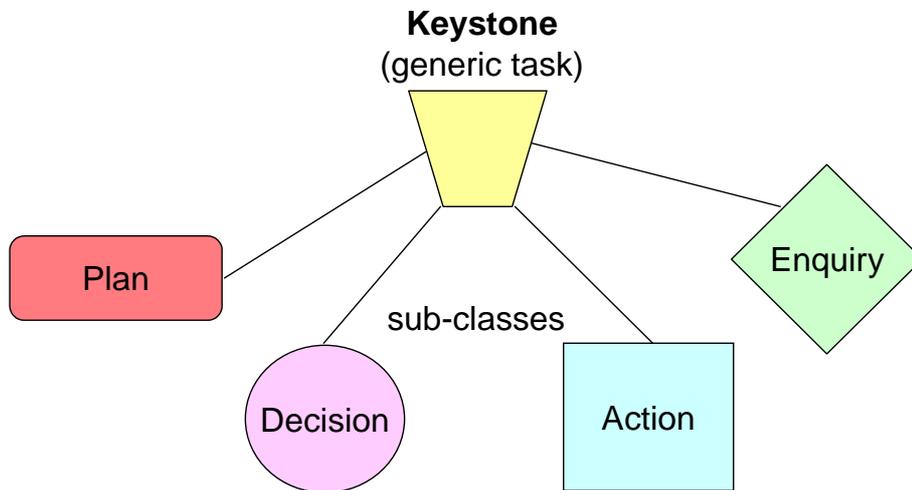


Figure 2.9 – Vue globale du modèle PROforma.

L'éditeur graphique Arezzo

L'environnement de développement *Arezzo*²⁷, consiste en un éditeur graphique permettant d'assister le processus de spécification d'un Guide de Bonnes Pratiques, puis de l'exécuter (Figure 2.10).

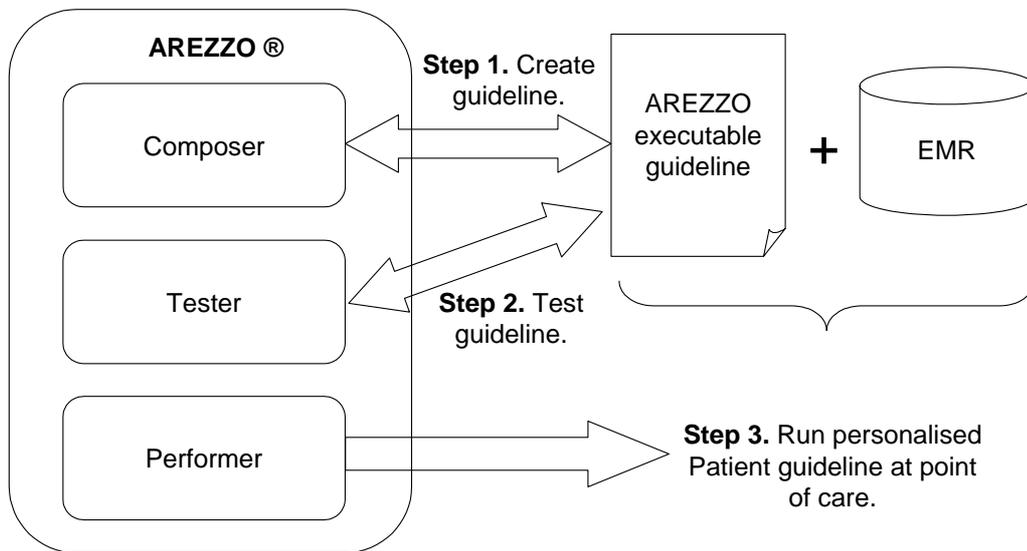


Figure 2.10 – Vue globale d'Arezzo²⁸.

²⁷ <http://www.infermed.com/arezzo> (dernier accès le 10 juin 2006).

²⁸ <http://www.infermed.com/arezzo/arezzo-components> (dernier accès le 10 juin 2006).

2.5.3. PROforma

Le *Composer* est utilisé pour créer les Guides de Bonnes Pratiques et les protocoles selon les quatre tâches décrites précédemment. Toutes connaissances procédurales ou médicales requises par le Guide de Bonnes Pratiques utilisent des motifs (*templates*) associés à chaque tâche, pour laquelle un état est défini, par exemple dormant (*Dormant*) signifie que la tâche n'a pas encore commencée ; requis (*Requested*) quand une tâche *requête* (*Enquiry*) est activée pour des valeurs spécifiques. L'état dormant (*Dormant*) correspond à l'état par défaut pour toutes les tâches avant que le Guide de Bonnes Pratiques ne s'exécute. Toutes les tâches contiennent également des attributs déterminant le comportement du Guide de Bonnes Pratiques pendant l'exécution du moteur *Performer* d'Arezzo. La structure graphique est instanciée (Figure 2.11), et est automatiquement convertie dans la base de données pour l'exécution.

Par exemple, pour l'extrait du Guide de Bonnes Pratiques de la prise en charge de la toux chronique, la représentation des recommandations dans PROforma correspond à la Figure 2.11. Le premier permet d'évaluer l'éligibilité du patient, pour lequel deux plans sont associés : celui correspondant au traitement initial et la radiographie (*CRX and initial treatment*) puis celui correspondant aux investigations (*Investigations*).

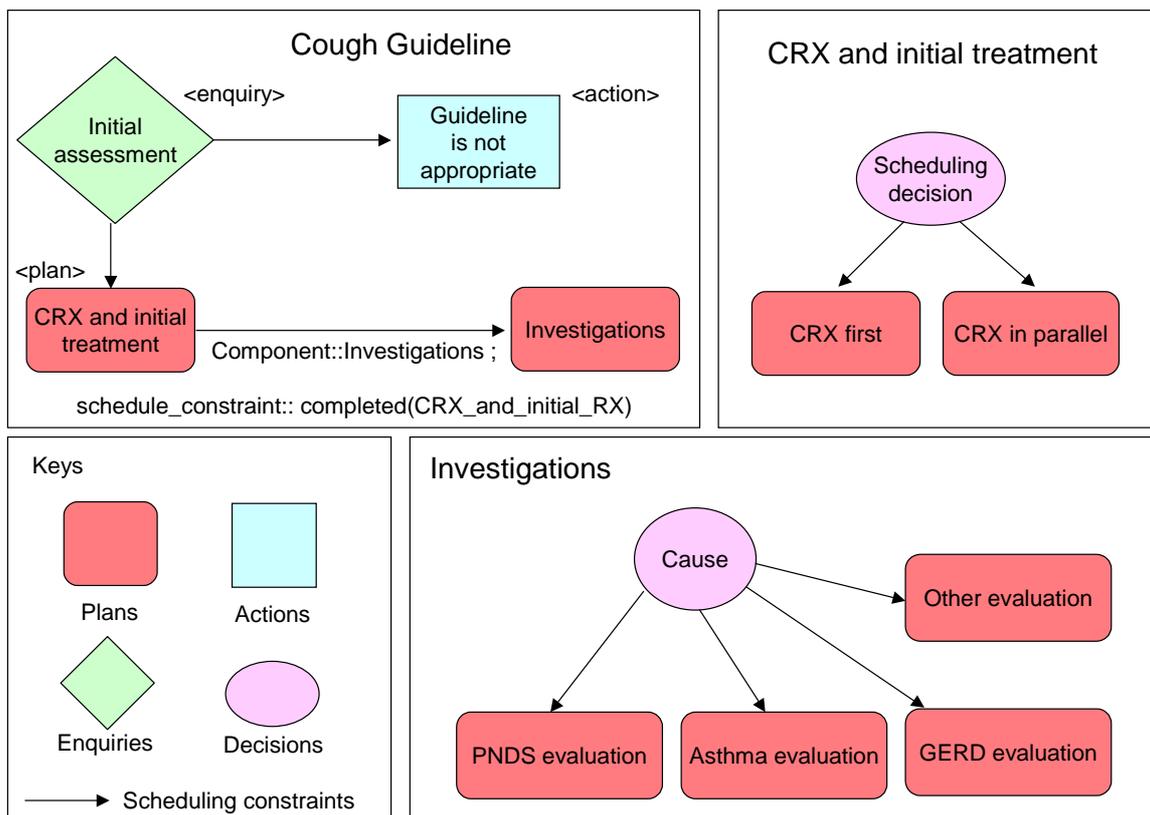


Figure 2.11 – Visualisation du Guide de Bonnes Pratiques de la toux chronique²⁹ en utilisant Arezzo.

²⁹ <http://smi-web.stanford.edu/auslese/smi-web/reports/SMI-2002-0923.pdf> (dernier accès le 10 juin 2006).

Selon le profil du patient, le deuxième niveau représente soit la radiographie en première intention soit la radiographie en parallèle d'un traitement (*CRX first* ou *CRX in parallel*). Le troisième niveau correspond aux investigations (*Investigations*) et sera exécutée dès que le traitement initial et la radiographie (*CRX and initial treatment*) ont été prescrits. Cet exemple montre la représentation des contraintes temporelles décrites dans le Guide de Bonnes Pratiques. Cette représentation permet de visualiser les grandes étapes de la conduite à tenir et, contrairement à GLIF, elle ne présente pas immédiatement les détails de la prise en charge.

Le Guide de Bonnes Pratiques créé avec le *Composer* est ensuite testé (*Tester* de la Figure 2.10) pour vérifier sa cohérence. Le moteur d'inférences *Performer* fournit les décisions cliniques selon le profil du patient, et permet d'intégrer les Guides de Bonnes Pratiques dans des systèmes existants, par exemple, en le couplant avec les dossiers médicaux informatisés (EMR des auteurs anglo-saxons).

En conclusion, *PROforma* offre un format déclaratif d'échange pour décrire les Guides de Bonnes Pratiques avec une méthode d'acquisition des connaissances et un ensemble d'outil pour simplifier la composition et la vérification formelle d'applications.

PROforma est la seule approche permettant de faire une distinction entre un langage déclaratif (par exemple, R^2L) utilisé pendant la phase d'acquisition des connaissances du Guide de Bonnes Pratiques et un langage procédural (par exemple, L_{R^2L}) traité par un interpréteur général pour l'exécution du moteur d'inférences (Fox et al., 2003).

Toutes les autres approches présentées dans ce chapitre demandent l'exécution d'un moteur d'inférences dédié, pour lequel les différents aspects procéduraux du Guide de Bonnes Pratiques sont encodés automatiquement. Les outils intégrés dans *PROforma* peuvent également être utilisés pour assister la normalisation des connaissances lors de l'élaboration des Guides de Bonnes Pratiques et des protocoles. Les Guides de Bonnes Pratiques encodés en *PROforma* sont également publiés sur le web en utilisant un logiciel dédié TALLIS³⁰.

2.5.4 GUIDE

L'approche GUIDE (Cicarese et al., 2003), développée à l'Université de Pavie fournit une infrastructure pour la gestion des connaissances médicales à travers un *workflow* appelé *careflow* (Quaglini et al., 2000; Quaglini et al., 2001).

Extension des réseaux de Petri

Comme précédemment, GUIDE cherche à modéliser les processus qui correspondent à la démarche diagnostique et thérapeutique des Guides de Bonnes Pratiques. Le formalisme utilisé dans GUIDE est basé sur les réseaux de Pétri, dans le but de modéliser des processus concurrents (séquentiel, parallèle, et itératif), souvent décrits dans les Guides de Bonnes Pratiques.

³⁰ <http://www.acl.icnet.uk/lab/tallis.html> (dernier accès le 10 juin 2006).

2.5.4. GUIDE

Les réseaux de Pétri sont étendus pour modéliser les aspects temporels (par exemple, pour estimer le temps dans chaque transition), les données (représentées en couleur) et les hiérarchies, décrits dans les Guides de Bonnes Pratiques. Un éditeur graphique permet d'assister la construction de réseaux de Pétri pour représenter les Guides de Bonnes Pratiques, comme celui de la prise en charge de la toux chronique (Figure 2.12).

En considérant l'exemple de la prise en charge de la toux chronique, on observe une succession d'étapes ordonnées, activées dès que l'étape précédente est validée. On peut observer une similitude avec GLIF du point de vue des détails de la prise en charge d'un patient. Une des différences concerne essentiellement la représentation des contraintes temporelles pouvant s'exécuter en parallèle et reflétant plus précisément la conduite à tenir décrite dans le Guide de Bonnes Pratiques. Par exemple, le fait de déterminer en parallèle si le patient est jeune non fumeur (*PNDS in young non-smokers*), ou si la femme est enceinte (*pregnant woman*), ou si il y a prise d'IEC (*patient on ACE inhibitors*).

L'approche Careflow

GUIDE est intégré dans un *workflow* considéré comme étant "a system that completely defines, manages, and executes workflow processes through execution of software whose order of execution is driven by a computer representation of the workflow process logic" (Quaglini et al., 2000). Le système permet d'implémenter entièrement un Guide de Bonnes Pratiques et de contrôler son exécution.

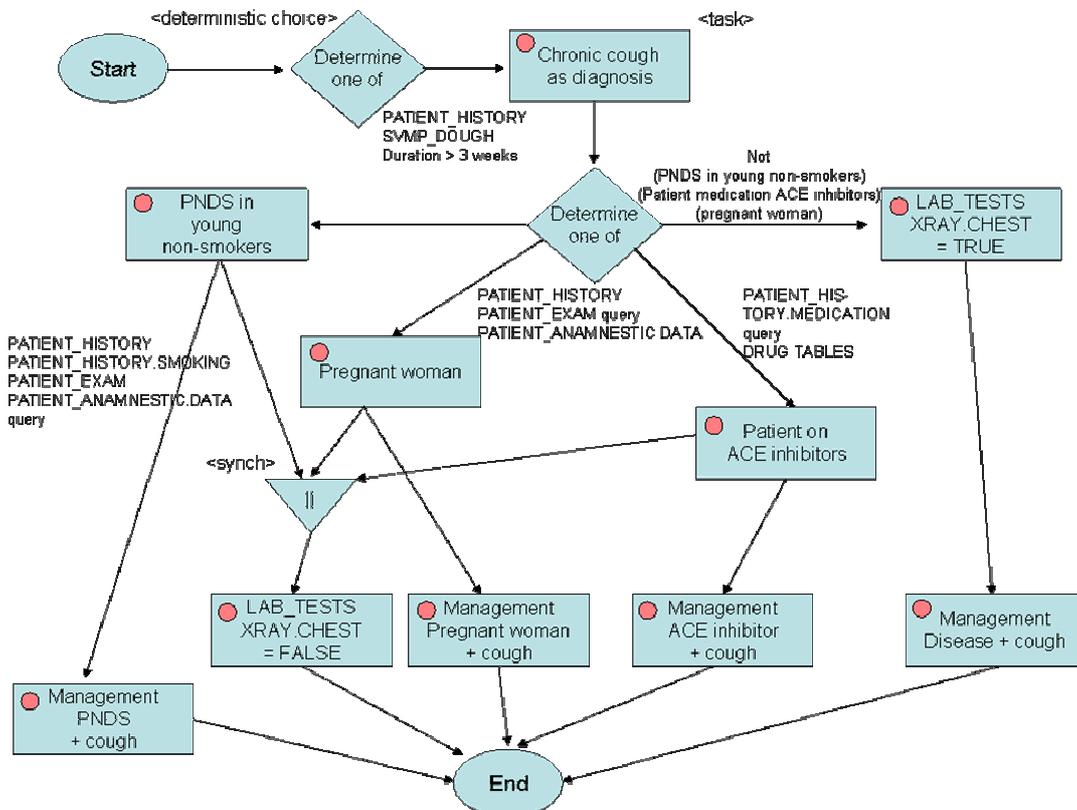


Figure 2.12 – Représentation du niveau supérieur (*top-level*) du modèle GUIDE pour l'extrait du traitement de la toux chronique (Georg, 2005).

Le “*Workflow Management Coalition*” définit un *workflow* comme : “*The automation of a business process, in a whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules*”. Cette définition peut être transférée au cas des processus cliniques décrits par les Guides de Bonnes Pratiques. L’équipe de GUIDE a introduit le terme *careflow* pour se rapporter au *workflow* dans le contexte de la prise en charge des patients. Quaglioni et ses collègues (Quaglioni et al., 2000) ont proposé une méthodologie permettant d’intégrer des outils de représentation des connaissances avec des outils commerciaux de *workflow* tels que Income™ et Oracle Workflow™, utilisés pour le modèle de simulation *careflow* et respectivement pour l’implémentation du *careflow*.

À partir des objets composant les réseaux de Pétri, une traduction dans un *Workflow Process Definition Language* (WPDL qui est le langage recommandé par le *Workflow Management Coalition*) est obtenue automatiquement. Les auteurs adoptent cette représentation standard dans le but d’exploiter différents produits existants pour les phases suivantes de l’implémentation du *Careflow*.

L’encodage dans le Careflow

La Figure 2.13 décrit les principales étapes permettant de construire un *careflow*. Les connaissances médicales et organisationnelles sont combinées pour construire un modèle computationnel du Guide de Bonnes Pratiques.

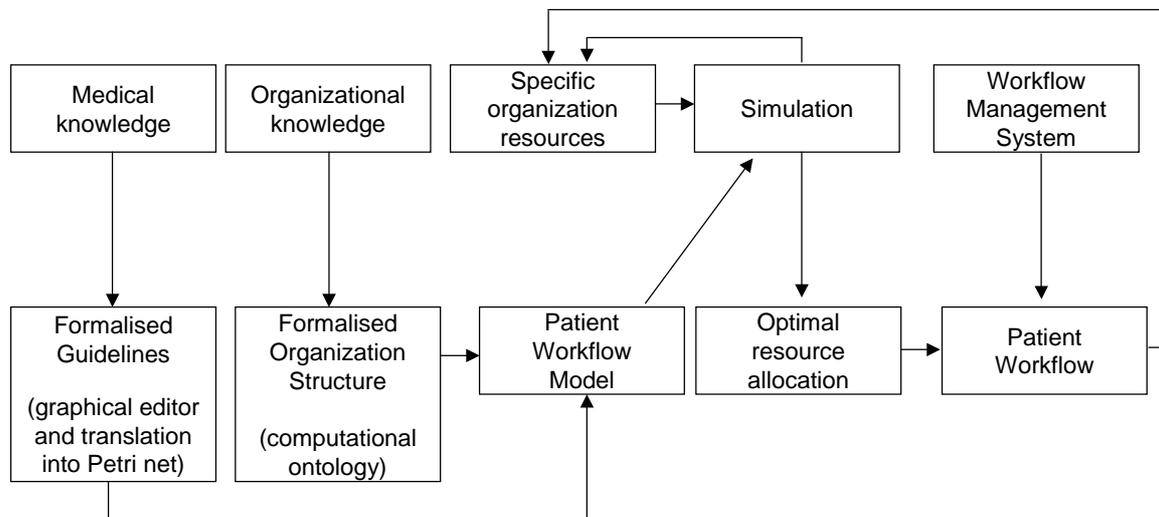


Figure 2.13 – Représentation de la méthodologie permettant de construire le *Careflow*.

GUIDE permet également de gérer des modules externes représentant des décisions comme des arbres de décision ou des diagrammes d’influence pouvant prendre en compte les préférences des patients, des préférences organisationnelles et des évaluations économiques.

Une nouvelle approche pour GUIDE

Récemment, l’environnement GUIDE a évolué en intégrant trois modules principaux : le *Guideline Management System* (fournissant un système d’aide à la décision), le *Healthcare Record System* (fournissant l’accès aux données patients), et le *Careflow Management System*

2.5.4. GUIDE

(fournissant les supports organisationnels) comme le montre la Figure 2.14 (Ciccarese et al., 2004). Cette nouvelle approche est basée sur le concept de séparation des connaissances médicales et organisationnelles (Ciccarese et al., 2005). Les Guides de Bonnes Pratiques peuvent alors être exécutés selon trois modes : (1) simuler des données patients ; (2) simuler l'implémentation de Guides de Bonnes Pratiques, rendue plus rapide par l'éditeur, et couplée à la base de connaissances reflétant la structure organisationnelle ; et (3) simuler la gestion des tâches selon les paramètres cliniques et l'utilisation (en utilisant l'outil Oracle™ Workflow).

En conclusion, des organismes de soins peuvent exécuter des Guides de Bonnes Pratiques avec l'éditeur GUIDE et les adapter à leurs besoins spécifiques. De nombreux projets basés sur cette approche sont actuellement en cours (Ciccarese et al., 2004; Ciccarese et al., 2005).

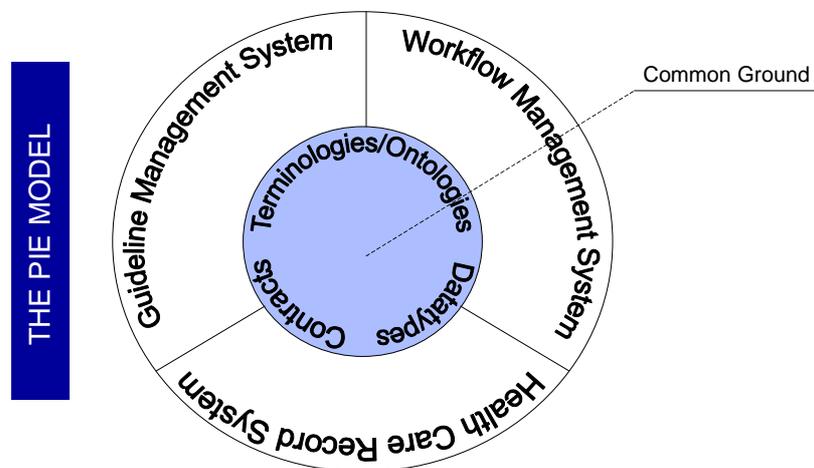


Figure 2.14 – Représentation du modèle Pie pour la séparation des trois modules principaux.

2.5.5 Aperçu des autres approches

Au-delà des principales approches présentées ci-dessus, il convient de mentionner d'autres techniques utilisées dans le domaine de l'informatisation des Guides de Bonnes Pratiques.

SAGE

Le système SAGE³¹ (*Standards-based sharable Active Guideline Environment*) (Campbell et al., 2003; Ram et al., 2004) commencé en 2002 est un projet impliquant IDX Corporation, Stanford Medical Informatics, Mayo Clinic (Rochester), l'Université du Nebraska, Intermountain Health Care, et Apelon Inc. Le but de ce projet consiste à proposer une architecture standard supportant tous les aspects du cycle de vie des Guides de Bonnes Pratiques, que ce soit lors de leur élaboration, de leur encodage, et de leur diffusion. SAGE, comme GLIF, a essentiellement pour vocation de partager et standardiser les Guides de Bonnes Pratiques suivant différentes institutions. Les cibles de ce projet concernent essentiellement les experts médicaux pour améliorer l'élaboration et l'encodage des Guides

³¹ <http://sageproject.net/> (dernier accès le 10 juin 2006).

de Bonnes Pratiques dans un format standard, ainsi que les organismes de soins pour diffuser les Guides de Bonnes Pratiques dans un format ne dépendant pas des systèmes d'informations cliniques. L'architecture de SAGE est représentée par la figure ci-dessous.

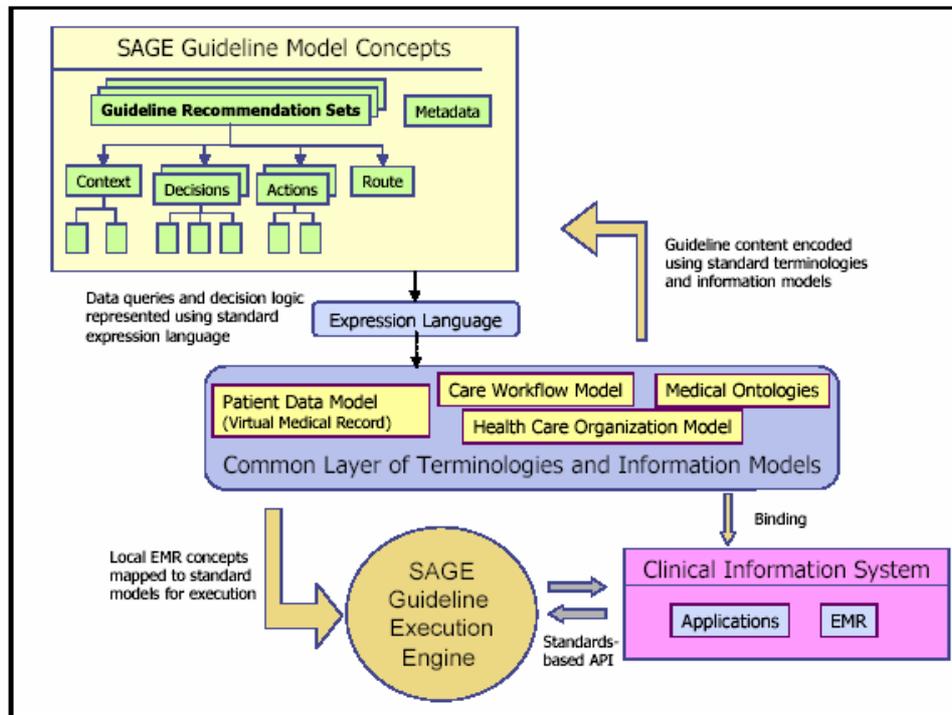


Figure 2.15 – Architecture de SAGE (Campbell et al., 2003).

Dans ce projet, l'encodage des Guides de Bonnes Pratiques s'inspire des techniques proposées par les autres approches, la nuance se situe essentiellement au niveau de la catégorisation des connaissances.

Des experts interprètent les Guides de Bonnes Pratiques et créent un ou plusieurs « plans » (on pourrait parler de scénarios, comme dans PRODIGY, décrit plus loin) qui correspondent aux objectifs du Guide de Bonnes Pratiques en fonction d'un contexte spécifique, par exemple l'organisation des soins. La recommandation se représente ensuite sous la forme de quatre nœuds :

- le contexte (*context*) permet de spécifier les évènements déclenchants du Guide de Bonnes Pratiques, comme les signes cliniques et l'état du patient ;
- la décision (*decision*) permet de décrire le processus d'acquisition de données (par exemple directement à partir du dossier patient informatisé) ;
- l'action (*action*) permet de déclencher les messages tels que les alertes *e-mail* ;
- la démarche (*route*) permet de synchroniser les différentes activités.

Ce projet a pour but d'élaborer des systèmes d'aide à la décision, couplés avec des dossiers patients informatisés qui permettent de fournir une recommandation au moment de la prise de décision clinique. SAGE permet de modéliser les Guides de Bonnes Pratiques complexes selon un ensemble de sous-recommandations, en utilisant des outils d'acquisition des connaissances comme Protégé (Musen et al., 1995).

PRODIGY

Le projet PRODIGY (*Prescribing Rationally with Decision-support in General-practice study*) (Purves et al., 1999) a pour but le développement et l'évaluation d'un système d'aide à la décision dédié aux médecins généralistes pour améliorer la qualité des soins par la diffusion de Guides de Bonnes Pratiques. Il permet également d'éditer des informations destinées au patient. PRODIGY a également pour but de fournir des « bonnes pratiques » quelle que soit la situation clinique du patient.

Les Guides de Bonnes Pratiques intégrés dans PRODIGY concernent essentiellement les prises en charge thérapeutiques de diverses pathologies aiguës et chroniques. Les phases I et II du projet ont permis de réaliser un module de consultation de Guides de Bonnes Pratiques couplé avec le dossier médical informatisé de certains fournisseurs industriels (respectant les standards du *National Health Service Information Authority*). PRODIGY (Figure 2.16) est actuellement largement utilisé en Angleterre par les médecins généralistes, et les Guides de Bonnes Pratiques qui y sont intégrés peuvent être consultés depuis un navigateur web³².

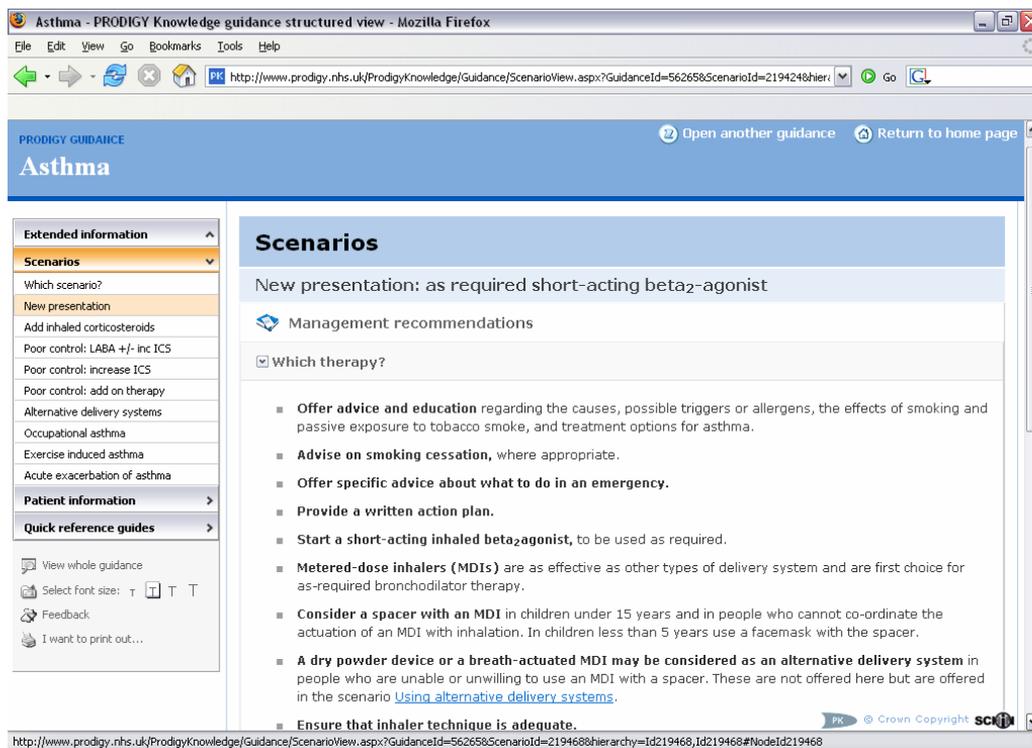


Figure 2.16 – Aperçu de Prodigy³² représentant le Guide de Bonnes Pratiques de l'asthme.

Les séquences décrites dans les Guides de Bonnes Pratiques sont représentées selon un réseau (Figure 2.17) consistant en des scénarios (*scenario*), des actions (*action steps*) et des sous-recommandations (*suguidelines*) qui permettent l'exécution d'autres scénarios (Johnson et al., 1999; Johnson et al., 2000). Ce modèle, influencé par le modèle EON (Tu & Musen, 2001; Wang et al., 2002b), introduit également la notion de scénarios mais ajoute la notion

³² <http://www.prodigy.nhs.uk/> (dernier accès le 10 juin 2006).

de consultation type (*consultation template*) qui représente généralement un contexte clinique pour lequel un scénario peut être identifié. Cette consultation type instancie les données du patient avec des valeurs pré-déterminées. L'objectif consiste à minimiser le temps nécessaire à la saisie de données par le clinicien (lorsque celles-ci n'ont pas été retrouvées dans le dossier médical informatisé ou lorsque le médecin veut simuler un cas patient).

Un scénario correspond donc à un état patient caractéristique, par exemple un diagnostic spécifique tel que « un hypertendu sans traitement médicamenteux ».

Pour chaque scénario, la réponse au traitement est considérée (*outcome assessments*) et une ou plusieurs actions (*action steps*) peuvent être suggérées au praticien. Les actions (représentant généralement la thérapie) sont associées à des états tels que initié (*initiate*), modifié (*modify*) ou arrêté (*stop*). Les scénarios sont également associés à un type de consultation (*consultation template*) décrivant les « bonnes pratiques » à suivre pour un patient donné, et proposent plusieurs actions alternatives selon la décision du médecin.

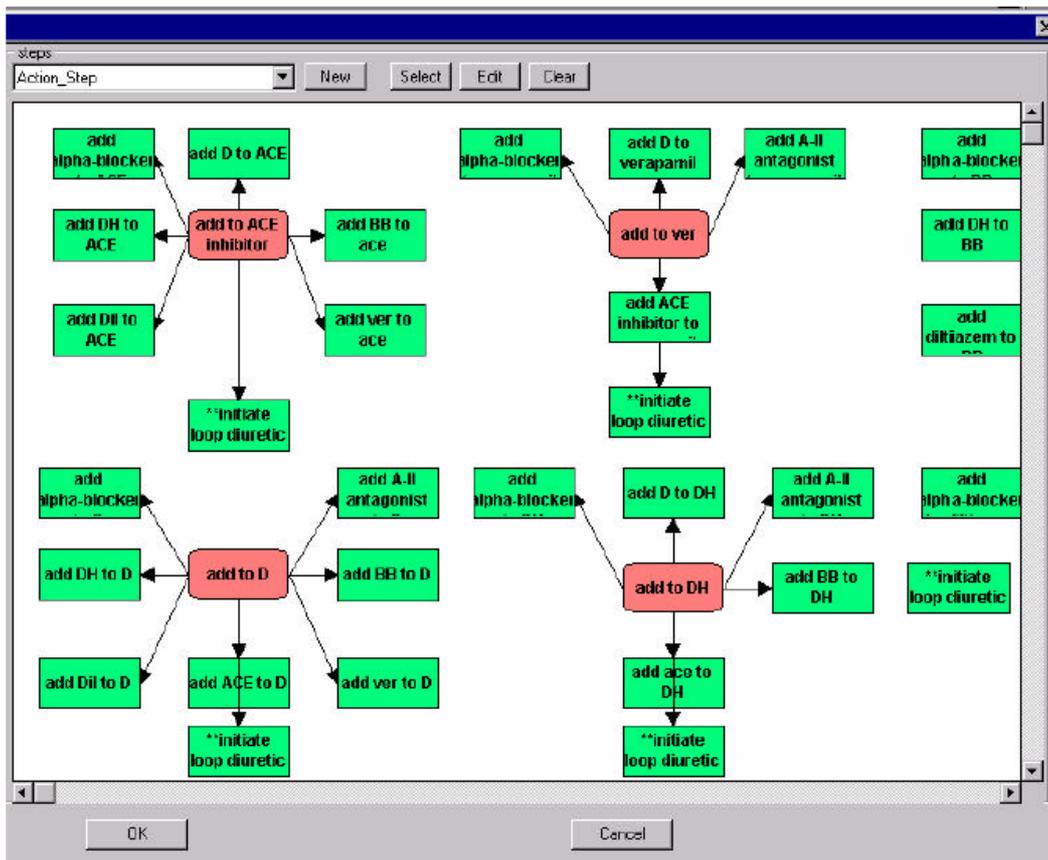


Figure 2.17 – Représentation de la prise en charge de la toux chronique composée de 4 scénarios.

ASBRU

Asbru³³ (Miksch, Shahar & Johnson, 1997) est un langage développé dans le cadre du projet Asgaard (Shahar, Miksch & Johnson, 1998) qui représente les Guides de Bonnes Pratiques selon des plans, en s'intéressant plus particulièrement aux aspects temporels. L'expression des contraintes temporelles dans Asbru permet de contrôler l'ordonnancement des tâches qui doivent être effectuées séquentiellement ou en parallèle lors de l'exécution d'un Guide de Bonnes Pratiques (Miksch et al., 1997). La syntaxe formelle est définie selon une forme BNF. Elle est proche de celle du langage LISP (Kosara & Miksch, 2001) mais n'est pas directement utilisable par des praticiens. AsbruView (Figure 2.18), un éditeur graphique dédié, a été développé permettant de visualiser et d'éditer les plans de traitement de Guides de Bonnes Pratiques (Miksch et al., 1998).

Asbru définit trois modes d'exécution d'un plan : séquentiel, concurrent et cyclique. Le modèle permet aussi de représenter de façon explicite la fréquence des actions qui doivent être exécutées de façon itérative jusqu'à ce que certaines conditions soient vérifiées. Les plans sont également composés des attributs suivants :

- les préférences (*preferences*) qui représentent l'applicabilité d'un plan (critères de sélection) et expriment son comportement ;
- les intentions (*intentions*) qui représentent les buts de haut niveau pouvant être atteints, maintenus ou évités durant l'exécution d'un plan ;
- les conditions (*conditions*) qui sont nécessaires à l'exécution d'un plan et pouvant avoir les états suivants : débuté (*started*), suspendu (*suspended*), réactivé (*reactivated*), abandonné (*aborted*), complété (*completed*) ;
- les effets (*effect*) qui décrivent les relations entre les arguments d'un plan et les paramètres mesurables par des fonctions mathématiques, par exemple pour les données médicales ;
- le corps du plan (*body plan*) qui décrit les plans ou les actions exécutées si les pré-conditions sont vérifiées.

Ce modèle comparable au modèle DILEMMA (Herbert et al., 1995) est donc basé sur une transition d'états pouvant simuler la mise en œuvre du plan en situation de soin. Le modèle Asbru définit quatre catégories d'intentions d'un plan :

- (1) l'état intermédiaire (*intermediate state*) qui représente l'état du patient devant être maintenu, atteint ou évité ;
- (2) l'action intermédiaire (*intermediate action*) qui représente l'action devant être effectuée lors de l'exécution d'un plan de soins ;
- (3) l'état global du motif (*overall state pattern*) qui représente l'état général du patient après que le plan de soins se soit correctement déroulé ;
- (4) l'action globale du motif (*overall action pattern*) qui représente l'action générale devant être entreprise.

³³ <http://smi-web.stanford.edu/projects/asgaard/AsbruL.html> (dernier accès le 10 juin 2006).

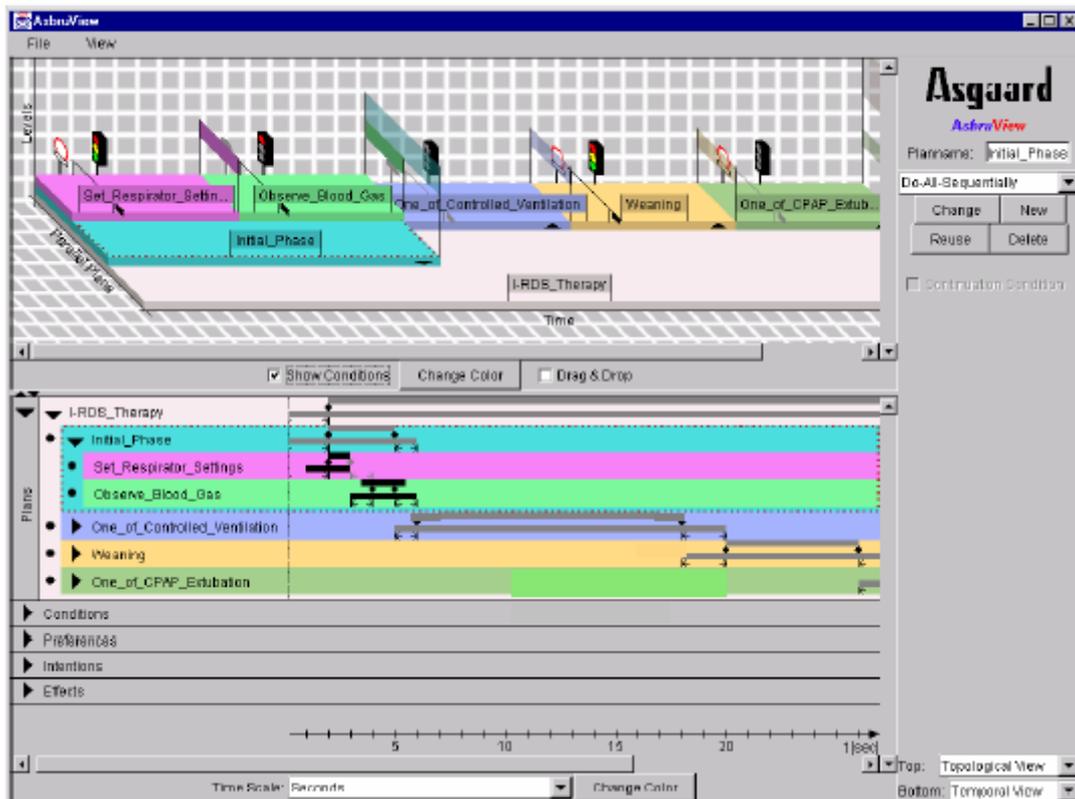


Figure 2.18 – Représentation du Guide de Bonnes Pratiques avec AsbruView.

GASTON

GASTON correspond à un *framework* (Figure 2.19) contenant plusieurs environnements dont le but consiste à faciliter les différentes étapes du cycle de vie des Guides de Bonnes Pratiques, de sa modélisation à son intégration dans un système d'aide à la décision (De Clercq, 2003).

GASTON contient donc :

- (i) un environnement permettant la modélisation et la visualisation des Guides de Bonnes Pratiques (Figure 2.20) de manière similaire à GLIF et EON ;
- (ii) un environnement de simulation permettant de tester et d'évaluer les Guides de Bonnes Pratiques développés de manière similaire à *PROforma* et *GUIDE* ;
- (iii) un environnement d'exécution pour coupler les Guides de Bonnes Pratiques (développés avec des systèmes d'informations patients) avec des bases de données médicales, de manière similaire aux approches actuelles et présentées précédemment.

2.5.5. Aperçu des autres approches

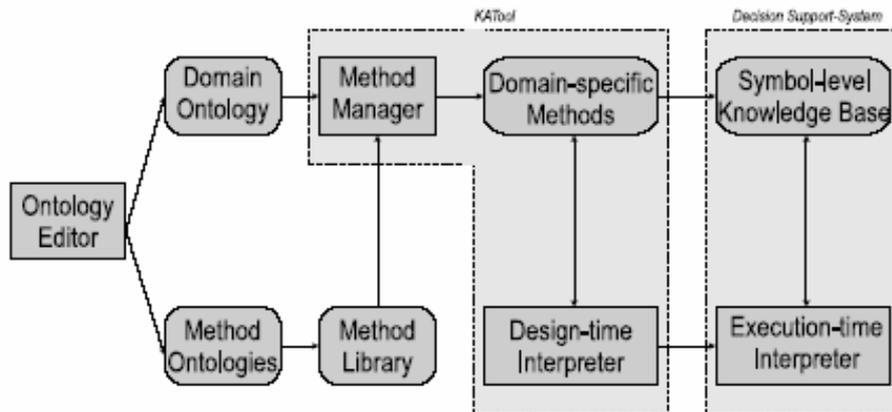


Figure 2.19 – Aperçu du *framework* de GASTON. Les rectangles correspondent à des programmes et les arrondis correspondent à des modèles, par exemple des ontologies (De Clercq et al., 2001).

Le formalisme de représentation des connaissances adopté dans GASTON, est basé sur les *frames*, et utilise les concepts des primitives de GLIF³⁴ (version 2), ainsi que des méthodes de résolution de problèmes, et des ontologies du domaine pour représenter le contenu des Guides de Bonnes Pratiques. Protégé³⁵ a également été utilisé pour faciliter la définition du domaine et le développement des ontologies. Celles-ci sont ensuite structurées selon CLIPS (Giarratano & Riley, 1994) pour faciliter leur intégration dans l'environnement d'exécution.

Concernant l'outil d'acquisition des connaissances, le KA-Tools (*Knowledge Acquisition Tools*) (Eriksson & Musen, 1993) a été à la base de GASTON pour faciliter le processus de modélisation des Guides de Bonnes Pratiques. Les experts peuvent donc formuler et structurer les connaissances du domaine, comparable à la philosophie de GUIDE (Ciccarese et al., 2003). Des algorithmes indépendants du domaine, basés sur les techniques de résolution de problèmes, sont mis en œuvre pour associer les rôles des connaissances contenues dans les Guides de Bonnes Pratiques aux connaissances du domaine.

L'approche GASTON, couplant plusieurs approches, définit des méthodes variées pour la détection d'erreurs dans les Guides de Bonnes Pratiques du fait des environnements dédiés représentant les différentes étapes du cycle de vie. Cette approche permet d'associer visuellement des informations lors de la phase d'acquisition des connaissances. Chaque classe de l'ontologie correspond également de manière automatique à un programme autonome permettant de tester « pas-à-pas » les connaissances encodées.

³⁴ Nous ne reviendrons pas dans cette partie sur la description des primitives de GLIF déjà décrites précédemment.

³⁵ Nous avons également fait mention de Protégé précédemment. <http://protege.stanford.edu/> (dernier accès le 10 juin 2006).

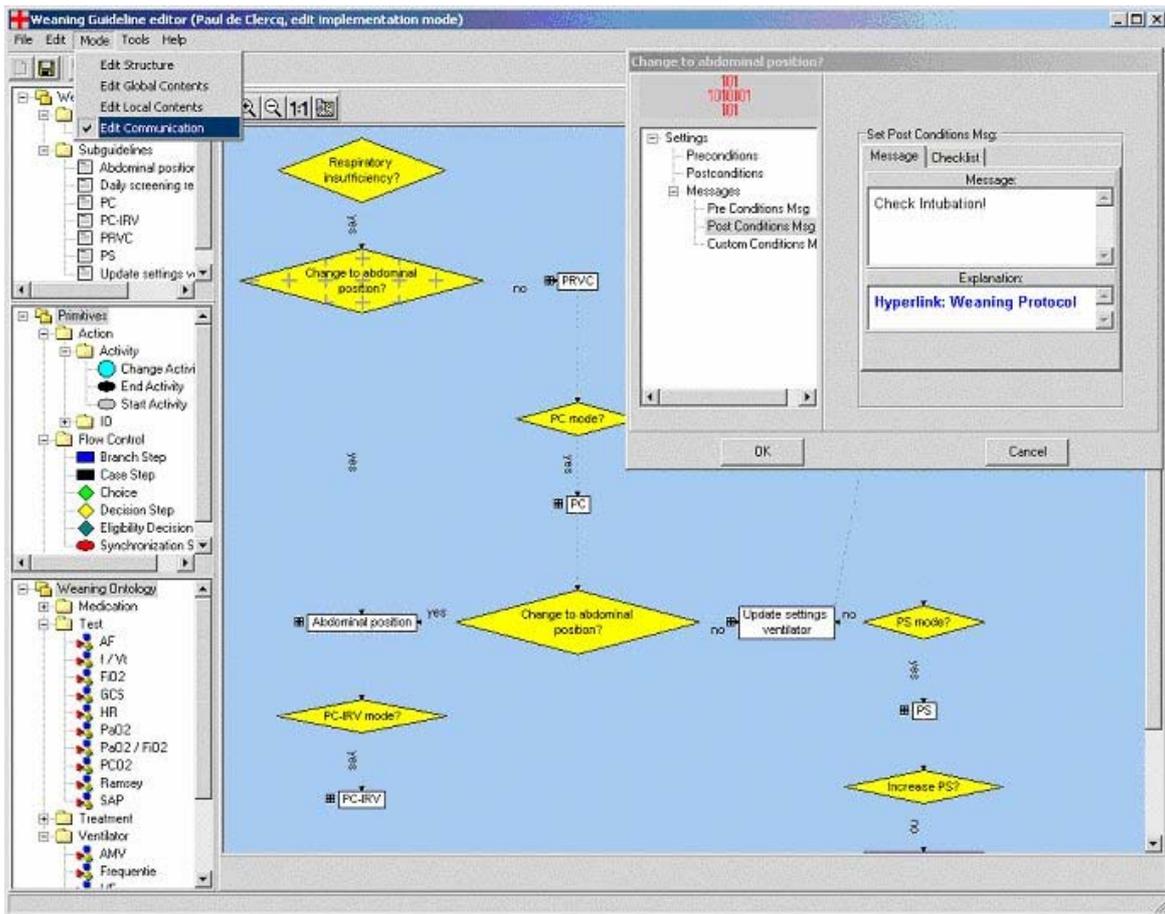


Figure 2.20 – Aperçu de l’environnement GASTON représentant la modélisation d’un Guide de Bonnes Pratiques³⁶.

Pour conclure cette section, nous présentons une petite synthèse (Tableau 2.4)³⁷ des caractéristiques de chaque approche « centrée sur les connaissances » présentées précédemment (De Clercq, 2003; Peleg et al., 2003; Terenziani et al., 2004; Wang et al., 2002a; Wang et al., 2001).

Il est important de souligner que les connaissances formalisées peuvent perdre de leur expressivité durant cette étape de formalisation (Van Der Maas, Ten Hoopen & Ter Hofstede, 2001). C’est une des raisons pour lesquelles des approches comme celles « centrées sur le document » ou *document-centered approach* se développent autour du texte lui-même plutôt qu’en lui substituant directement des formalismes de représentation. Pour ces approches, une meilleure standardisation des Guides de Bonnes Pratiques permettrait de contrôler le contenu du document et faciliter ainsi le passage du texte vers des formats computationnels.

³⁶ http://www.openclinical.org/gmm_gaston.html (dernier accès le 10 juin 2006).

³⁷ Complétée par celles manquantes telles que SAGE, GLARE, GASTON.

2.5.5. Aperçu des autres approches

Tableau 2.4 – Caractéristiques des modèles pour représenter les connaissances contenues dans les Guides de Bonnes Pratiques.

Modèle de représentation	Primitives des représentations				Structure des primitives		Modélisation des données patient
	Décision	Action	État du Patient	Exécution	Contraintes temporelles	Imbrication	
Syntaxe Arden	propriété logique (<i>logic slot</i>)	action (<i>action slot</i>)	aucune (<i>no</i>)	aucune (<i>no</i>)	module d'invocation (<i>module invocation</i>)	aucune (<i>no</i>)	aucune (<i>no</i>)
DILEMMA / PRESTIGE	transition (<i>state transition</i>)	protocole (<i>protocol</i>)	n/a	procédure (<i>procedure state</i>)	protocole (<i>protocol state</i>)	protocole, composition, diagramme de transition (<i>protocol composition, state transition diagram</i>)	modèle de dossiers patient (<i>patient record model</i>)
EON/ DHARMA	décision (<i>decision step</i>)	action, activité (<i>action, activity</i>)	scénario, activité (<i>scenario, activity state</i>)	aucune* (<i>no*</i>)	organigramme (<i>flowchart</i>)	sous-ensemble de recommandations (<i>subguideline</i>)	ontologies du domaine pour le dossier médical (<i>EPR ontology</i>)
PROforma	décision (<i>decision</i>)	action, requête (<i>action, enquiry</i>)	n/a	tâches (<i>task state</i>)	graphe des contraintes (<i>constraints satisfaction graph</i>)	plan (<i>plan</i>)	n/a
GLIF	décision (<i>decision step</i>)	action (<i>action step</i>)	état du patient (<i>patient state step</i>)	aucune* (<i>no*</i>)	diagramme (<i>flowchart</i>)	sous-ensemble de recommandations (<i>subguideline</i>)	ontologie du domaine à trois niveaux (<i>three-layer domain ontology</i>)
Asbru	condition, préférences (<i>condition, preference</i>)	plan (<i>plan</i>)	motifs temporels (<i>temporal patterns</i>)	plan (<i>plan state</i>)	corps du plan (<i>plan-body</i>)	plan (<i>plan</i>)	n/a
GUIDE / Patman	décision (<i>decision</i>)	tâche, attente, surveillance (<i>task, wait, monitor</i>)	(implicite) (<i>implicit</i>)	n/a	diagramme (<i>flowchart</i>)	tâche (<i>task</i>)	modèle relationnel (<i>relational</i>)

PRODIGY	décision (<i>decision</i>)	action, activité (<i>action, activity</i>)	scénario (<i>scenario</i>)	n/a	diagramme de transitions (<i>state transition diagram</i>)	sous-ensemble de recommandations (<i>subguideline</i>)	ontologies du domaine pour le dossier médical (<i>EPR ontology</i>)
SAGE	décision (<i>decision</i>)	actions (<i>actions</i>)	contexte (<i>context</i>)	démarche (<i>route</i>)	diagramme (<i>flowchart</i>)	sous-ensemble de recommandations (<i>subguideline</i>)	modèle de données patient / ontologies médicales (<i>patient data model / medical ontologies</i>)
GLARE	décision (<i>decision atomic</i>)	action d'exécution, action d'information (<i>work actions, query actions atomic</i>)	conclusions (<i>conclusions atomic</i>)	n/a	graphe de contraintes (<i>constraints satisfaction graph</i>)	action complexe (<i>plan composite action plan</i>)	base de données patient (<i>patient database</i>)
GASTON	décision (<i>decision</i>)	action (<i>action</i>)	n/a	aucune* (<i>no*</i>)	diagramme de contraintes (<i>constraints satisfaction flowchart</i>)	sous-ensemble de recommandations (<i>subguideline</i>)	systèmes d'information patients, base de données médicales (<i>patient information systems, medical databases</i>)

n/a : informations non disponibles à partir des publications

* EON/DHARMA, GLIF et GASTON (basé sur GLIF) contiennent des états d'exécution (*execution states*), mais il n'y a pas de modèle de représentation pour l'exécution du Guide de Bonnes Pratiques.

2.6 Approches documentaires

On désigne sous le terme d'approche documentaire les méthodes qui utilisent le document lui-même comme support de représentation. Ces méthodes ont été rendues possibles avec le développement de langages de balisage comme XML (introduit dans le chapitre 4). Les Guides de Bonnes Pratiques encodés en XML peuvent être structurés, enrichis de multiples connaissances et annotations, tout en pouvant être visualisés dans différents formats, en fonction des besoins des utilisateurs, et le tout, en préservant leur forme originale de publication (Hagerty et al., 2000). Le balisage des Guides de Bonnes Pratiques introduit la capacité d'imposer une structure tout en définissant les relations entre les données et en y ajoutant une capacité à définir ses propres éléments. Cette approche s'est avérée efficace pour fournir une intégration plus simple et moins dépendante des processus ou des *workflows* dans lesquels les Guides de Bonnes Pratiques s'intègrent.

2.6.1 Hypertext Guideline Markup Language

Hypertext Guideline Markup Language ou HGML (Hagerty et al., 2000) permet de représenter des Guides de Bonnes Pratiques à l'aide de balises XML pour marquer le contenu du document et d'autres données pertinentes (méta-données), et faciliter la conversion du texte vers un format informatique (Hagerty et al., 2000). Un éditeur dédié (Figure 2.21) permet de baliser manuellement le contenu du Guide de Bonnes Pratiques. En effet, il est plus simple pour l'utilisateur non familier des langages informatiques de sélectionner les connaissances à marquer dans un éditeur générant automatiquement le fichier XML.

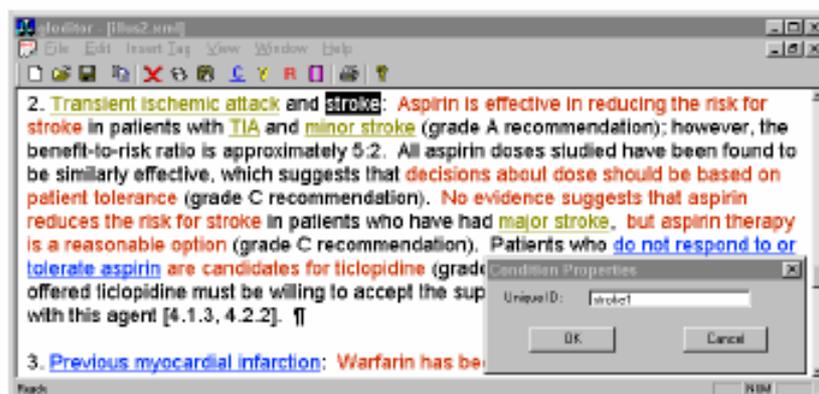


Figure 2.21 – Éditeur facilitant le marquage du Guide de Bonnes Pratiques (Hagerty et al., 2000).

Des balises sont donc disponibles (Figure 2.22) pour faciliter les inférences du contenu des Guides de Bonnes Pratiques. (Hagerty et al., 2000).

Les axes de recherche en cours concernent deux aspects, l'un permet l'identification de termes médicaux grâce à la terminologie et les concepts d'UMLS³⁸, l'autre permet de rechercher certains termes des Guides de Bonnes Pratiques en utilisant l'environnement de

³⁸ <http://www.nlm.nih.gov/research/umls/> (dernier accès le 10 juin 2006).

traitement automatique des documents, GATE³⁹ (développé à l'Université de Sheffield), qui a été mentionné comme exemple pour identifier automatiquement des expressions conditionnelles (Hagerty et al., 2004).

```
<statement> <recommendation> <keyword> Aspirin </keyword> is
effective in reducing the risk for stroke </recommendation> <condition> in patients with
<condition> TIA </condition> and <condition> minor stroke </condition> </condition>
</statement>
```

Figure 2.22 – Représentation d'une recommandation balisée en HGML (Hagerty et al., 2000).

En résumé, HGML permet le balisage manuel des Guides de Bonnes Pratiques par des experts à l'aide d'un éditeur dédié, dans le but de produire une version structurée du document. Les premiers résultats du balisage HGML manuel montrent qu'un peu moins de 61% des recommandations contenues dans le document sont identifiées. Les travaux futurs consistent à développer des outils d'encodage de Guides de Bonnes Pratiques basés sur des techniques d'apprentissage. Cette approche a également l'ambition de pouvoir normaliser automatiquement les termes contenus dans les Guides de Bonnes Pratiques, et de construire des bases de connaissances couplées avec un dossier médical informatisé.

On peut comparer brièvement la démarche d'encodage d'un Guide de Bonnes Pratiques dans une approche documentaire à celle d'une approche à base de connaissances. Dans le second cas, il s'agit de façon générale, d'une démarche d'acquisition de connaissances à partir du texte. Les formalismes représentent les différents processus avec leurs décompositions, leurs aspects temporels et conditionnels. C'est pourquoi, ils se répartissent entre représentations algorithmiques, plans et réseaux de Pétri. Le document doit être analysé et compris dans son ensemble pour pouvoir être traduit dans ces formalismes. L'approche documentaire, quant à elle, enrichit le texte d'éléments d'informations : elle peut découler d'une analyse locale et progressive au fil du document. Son niveau de représentation peut paraître moins ambitieux, néanmoins elle peut constituer une représentation pivot servant de point de départ à des transformations ultérieures ou à la génération d'autres représentations, comme des règles de décision.

2.6.2 Guideline Elements Model (GEM)

L'approche documentaire la plus connue est *Guideline Elements Model* (GEM) (Shiffman et al., 2000) développé par l'équipe de Yale, et approuvé comme standard ASTM E2210-02. Le modèle est basé sur une DTD XML et a pour but de structurer les connaissances hétérogènes contenues dans les Guides de Bonnes Pratiques, et faciliter le passage du contenu des Guides de Bonnes Pratiques vers un format informatique.

Une centaine d'éléments a été définie pour catégoriser chaque connaissance contenue dans les Guides de Bonnes Pratiques : ces éléments sont représentés selon une hiérarchie composée de neuf branches principales comme le montre la Figure 2.23. Les neuf branches correspondent aux éléments suivants :

³⁹ <http://gate.ac.uk/> (dernier accès le 10 juin 2006).

2.6. Approches documentaires

- L'identité (*Identity*) contient les paramètres d'identification des Guides de Bonnes Pratiques tels que le titre, la référence et la date de publication, le format, le contact, etc.
- Le développeur (*Developer*) permet de décrire l'organisation à l'origine de la rédaction du document.
- L'objectif (*Purpose*) décrit les objectifs de développement du Guide de Bonnes Pratiques (diagnostic, thérapeutique ou prévention), ainsi que, par exemple, les indicateurs des retombées sanitaires.
- L'effet prévu (*Intended Audience*) concerne les médecins dont le Guide de Bonnes Pratiques est supposé modifier le comportement.
- La méthode de développement (*Method of Development*) concerne la méthodologie sous-jacente à l'élaboration du Guide de Bonnes Pratiques (méta-analyses, niveaux de preuves, etc.).
- La population cible (*Target Population*) permet de caractériser la population des patients concernés.
- La relecture (*Testing*) concerne le comité de relecture du Guide de Bonnes Pratiques (différent du groupe de travail en charge de la rédaction).
- Le plan d'actualisation (*Review Plan*) contient les dates des prochaines révisions permettant de délimiter dans le temps la validité du Guide de Bonnes Pratiques.

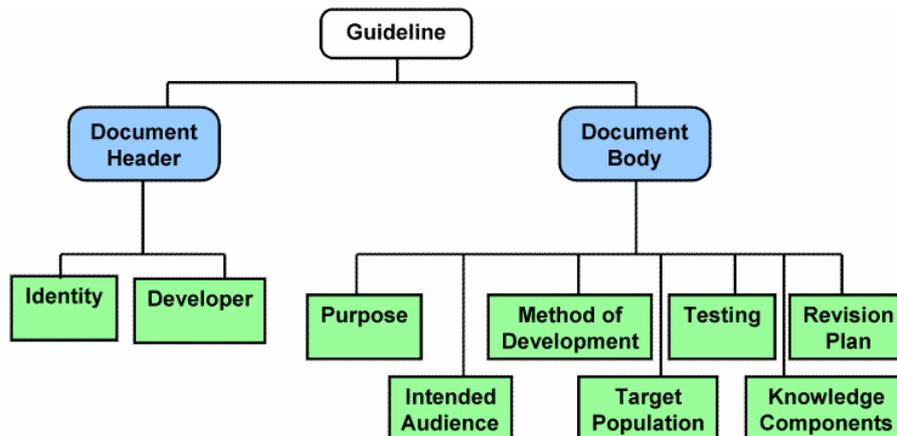


Figure 2.23 – Représentation de la hiérarchie GEM.

- Les composants de connaissances (*Knowledge Components*) (Figure 2.24) permettent de structurer les recommandations du point de vue des actions préconisées.

Plusieurs balises constituent les composants de connaissances (*Knowledge Components*) pour lesquelles des termes spécifiques ont été définis par le *National Clearinghouse*⁴⁰, une source de vocabulaire contrôlé, permettant ainsi de mieux structurer le texte essentiellement pour son informatisation (Bernstam et al., 2000).

Outre les éléments permettant la caractérisation de la méthodologie à l'origine de l'élaboration du Guide de Bonnes Pratiques (description du groupe de travail, recherche documentaire dans les bases de données bibliographiques, méta-analyses, calcul du niveau

⁴⁰ <http://www.guideline.gov> (dernier accès le 10 juin 2006).

de preuve), on s'intéresse en particulier à l'élément qui s'attache à structurer les recommandations du point de vue des actions préconisées (*Knowledge Components*). Les recommandations peuvent être de deux types : (i) impératives (*imperative*) lorsqu'elles s'appliquent à l'ensemble de la population cible du Guide de Bonnes Pratiques, (ii) conditionnelles (*conditional*) quand elles s'appliquent à une catégorie spécifique d'individus identifiée par des critères particuliers.

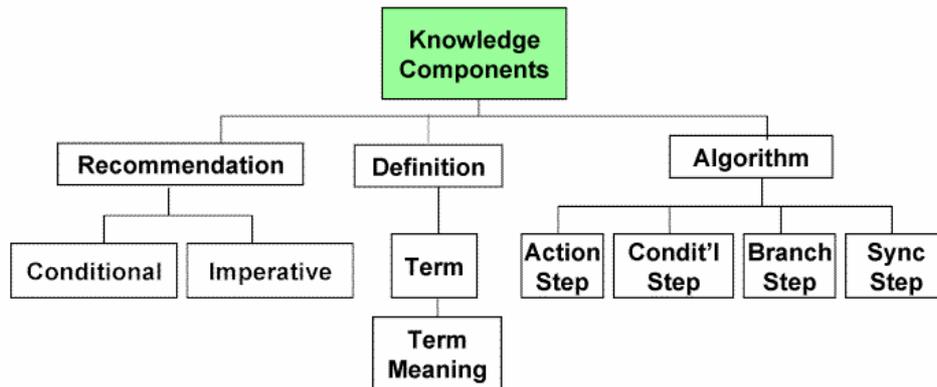


Figure 2.24 – Représentation de la hiérarchie pour l'élément *Knowledge Components*.

Les termes imprécis utilisés dans les recommandations sont définis dans l'élément définition (*definition*). L'enchaînement des recommandations les unes par rapport aux autres dans un plan de soin global est décrit dans l'élément algorithmique (*Algorithm*), sous la forme d'une séquence temporellement ordonnée d'actions (basée sur le modèle GLIF).

GEM est attendu pour faciliter le passage du langage naturel des Guides de Bonnes Pratiques dans un format standard et informatisable. L'encodage GEM des connaissances des Guides de Bonnes Pratiques est possible à travers le processus de balisage qui ne demande pas de compétences informatiques particulières. L'application GEM-Cutter⁴¹ est un éditeur XML qui facilite le balisage interactif des Guides de Bonnes Pratiques⁴² (Gershkovich & Shiffman, 2001) en évitant à l'utilisateur d'avoir à inscrire « physiquement » des balises dans le texte.

GEM est aussi destiné à être utilisé à travers les différentes étapes du cycle de vie d'informatisation des Guides de Bonnes Pratiques, depuis leur écriture (Shiffman, Michel & Essaihi, 2004a) jusqu'à leur implémentation dans des systèmes d'aide à la décision (Shiffman et al., 2004b). Néanmoins, le balisage des Guides de Bonnes Pratiques présente plusieurs limites. Bien qu'il ait été désigné comme un modèle complet pour modéliser le contenu des Guides de Bonnes Pratiques, des variations significatives sont observées lors de sa mise en œuvre et de la création des documents GEM à partir d'un Guide de Bonnes Pratiques par différents intervenants (Karras, Nath & Shiffman, 2000). Comme ce modèle est une simple abstraction des Guides de Bonnes Pratiques, GEM seul ne peut pas supporter la résolution des ambiguïtés présentes dans différents Guides de Bonnes Pratiques textuels.

⁴¹ <http://gem.med.yale.edu> (dernier accès le 10 juin 2006).

⁴² Nous décrivons dans la section suivante les caractéristiques de cette interface.

Face aux problèmes rencontrés, Shiffman (Shiffman et al., 2004a) a récemment introduit des étapes pour interpréter le texte à encoder :

- L'atomisation (*atomization*) consiste à extraire et spécifier des concepts élémentaires tels que l'identification des variables de décision et des actions, et la traduction dans un modèle de données (ou une terminologie contrôlée). Cette catégorisation permet d'affiner le niveau de représentation par rapport au texte.
- La désabstraction (*deabstraction*) consiste à ajuster le niveau d'expression du Guide de Bonnes Pratiques à un niveau permettant l'opérationnalisation.
- La désambiguïsation (*disambiguation*) permet de préciser les variables de décision et les actions afin qu'il n'y ait qu'une seule interprétation possible et rétablir l'exclusivité des modalités.
- La complétude (*verification of completeness*) consiste à ajouter des notions médicales afin qu'il y ait des recommandations pour toutes les situations cliniques possibles, par exemple s'assurer de l'exhaustivité de la caractérisation de chaque situation clinique, et rétablir l'exhaustivité de l'ensemble des situations cliniques possibles.

2.7 Interfaces dédiées aux approches documentaires

Si les approches documentaires demeurent proches du texte, la manipulation directe du texte enrichi ou balisé n'en est pas moins difficilement acceptable par les utilisateurs. Ce problème rejoint donc, paradoxalement, celui de l'encodage des Guides de Bonnes Pratiques dans des formalismes informatiques. Dans cette section, nous nous intéressons aux différentes interfaces développées dans le cadre des approches documentaires dont le but consiste essentiellement à faciliter le passage du texte des Guides de Bonnes Pratiques vers des modèles tel que GEM, ou vers des langages tel que Asbru.

2.7.1 GEM-Cutter

GEM-Cutter⁴³ (Figure 2.25) a été développé par l'équipe de Yale dans le but d'assister le processus de balisage manuel du texte des Guides de Bonnes Pratiques. GEM-Cutter a été défini comme un éditeur XML, et est essentiellement une interface conviviale (*'user-friendly'*) pour la manipulation de texte, le balisage des sélections textuelles et l'encodage du document. Il contient une arborescence complète des concepts de GEM qui facilite l'accès à chaque catégorie (partie centrale de la Figure 2.25). L'ensemble de l'interface permet donc la manipulation interactive du document. Même si GEM-Cutter ne permet pas l'identification automatique du *contenu* du document, il permet néanmoins de diminuer considérablement la charge cognitive de l'utilisateur en lui offrant directement les informations relatives aux catégories de GEM et surtout en supportant le processus incrémental du balisage du document. L'interface GEM-Cutter est basée sur un système de

⁴³ <http://gem.med.yale.edu/> (dernier accès le 10 juin 2006).

fenêtres multiples qui affichent le texte original du Guide de Bonnes Pratiques (partie gauche de la Figure 2.25), l'explorateur des catégories de GEM et le texte correspondant encodé en GEM (partie droite de la Figure 2.25). L'interface fonctionne de la façon suivante : on sélectionne une portion de texte et on la déplace vers une catégorie de GEM, ce qui génère une entrée correspondante dans le document encodé en GEM, consistant en une portion de texte balisée par la catégorie dans laquelle elle a été déplacée.

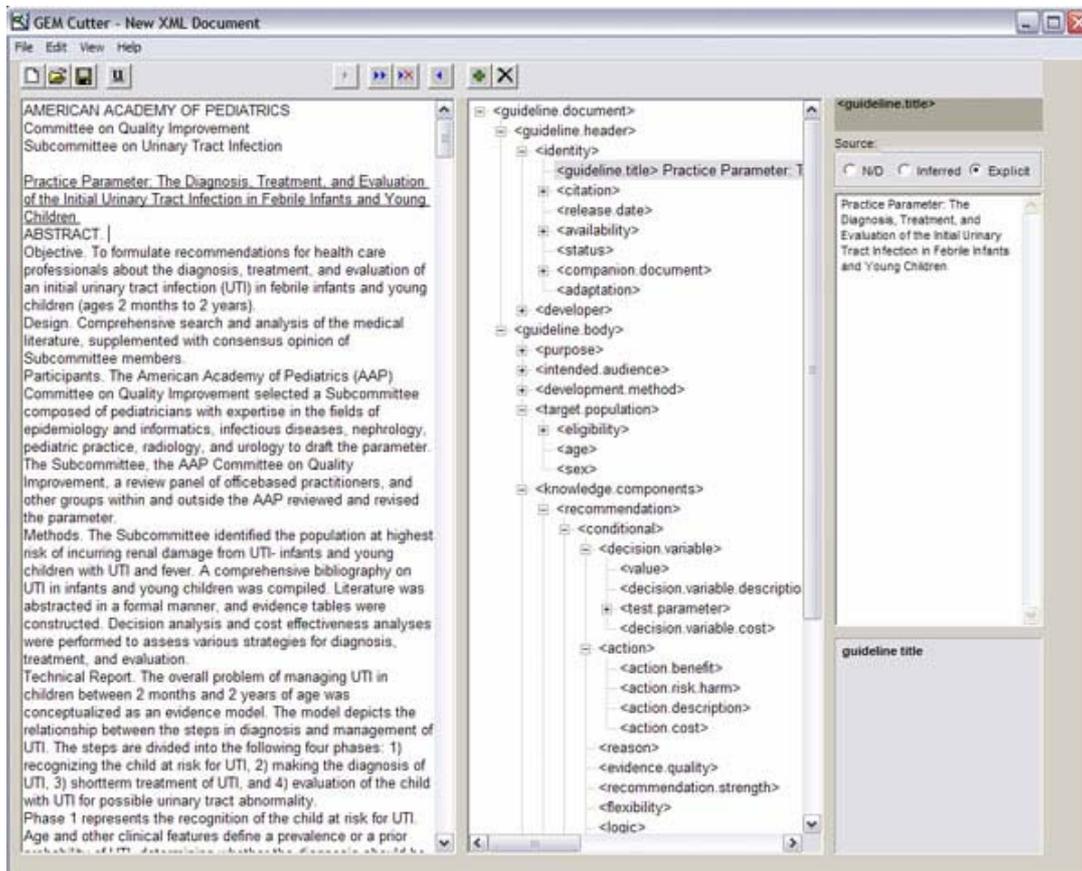


Figure 2.25 – Interface représentant l'éditeur GEM-Cutter⁴⁴ dédié à l'encodage des Guides de Bonnes Pratiques selon le modèle GEM.

2.7.2 Guide-X / Stepper

Dans une approche similaire à GEM-Cutter, Stepper (Figure 2.26) (Svatek & Ruzicka, 2003) supporte l'encodage des Guides de Bonnes Pratiques à travers un processus de balisage du texte par une succession d'étapes, appelée *step-by-step*, jusqu'à la génération finale du document formalisé. Cette interface a pour vocation de minimiser la perte d'informations pendant le processus d'encodage « *minimize information loss during the encoding process*. ». La principale différence par rapport à GEM-Cutter concerne le nombre d'étapes, une seule pour GEM-Cutter alors que Stepper propose une décomposition explicite des différentes phases de formalisation du document, pour lesquelles cinq étapes ont été

⁴⁴ <http://gem.med.yale.edu/> (dernier accès le 10 juin 2006).

2.7. Interfaces dédiées aux approches documentaires

identifiées comme nécessaires pour passer entièrement du texte vers des connaissances opérationnelles.

Stepper, basé sur XML, contient un processeur XSLT⁴⁵ (*eXtensible Stylesheet Language Transformation*) permettant de supporter les différentes étapes de balisage et de générer le document formalisé. Les transformations sont gérées par des règles exprimées dans un nouveau langage de transformation basé sur XML appelé XKBT (*XML Knowledge Block Transformation*). Les transformations des premières étapes ont un caractère plutôt d'amélioration du balisage, alors que les étapes ultérieures consistent en l'agrégation des buts et des scénarios.

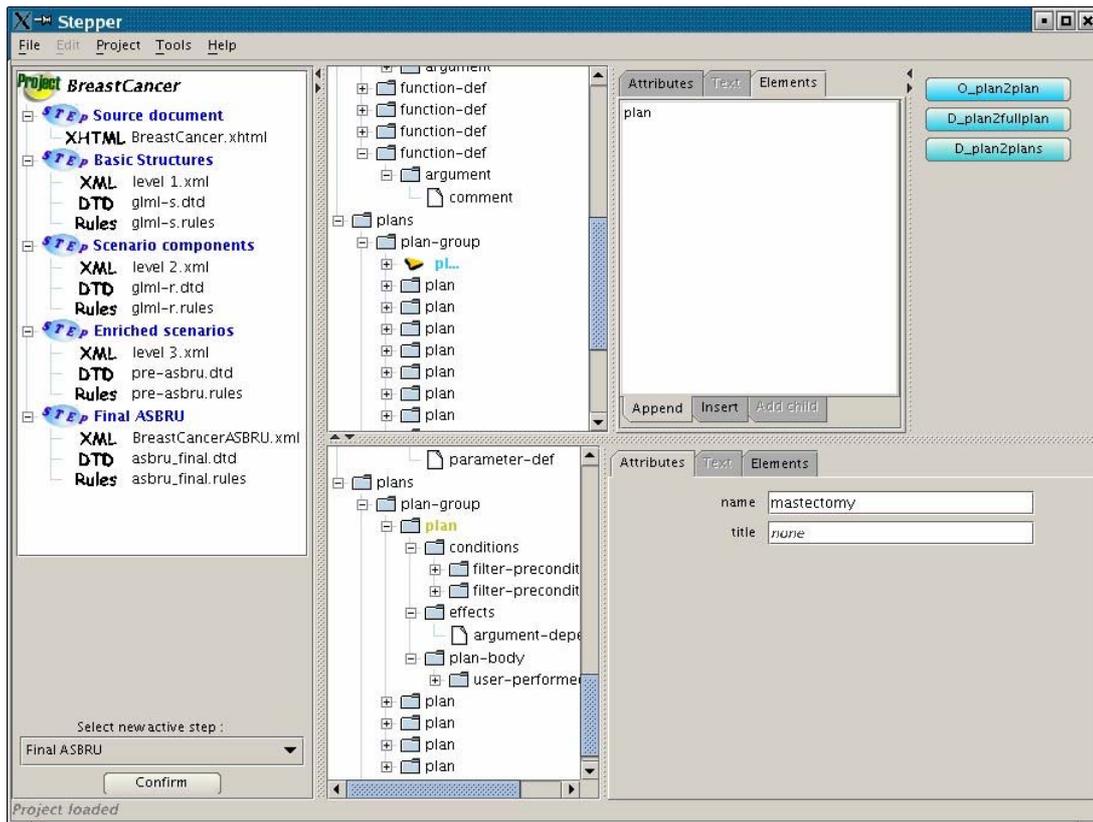


Figure 2.26 – Interface de Stepper représentant l'encodage d'un Guide de Bonnes Pratiques⁴⁶.

Dans l'environnement Stepper, tout le document est parcouru, et les segments sélectionnés sont assignés à quatre catégories : composant procédural (*procedural component*) (notion de scénario), de causalité (*causality*), de but (*goal statement*), et de définition de concepts (*concept definition*). En fonction des catégories, des règles XKBT sont déclenchées. L'élément procédural (*procedural*), inclut dans une structure de scénario, représente la condition d'une recommandation.

L'évaluation a montré que l'approche *step-by-step* rend possible la transformation de Guides de Bonnes Pratiques en fragments de codes opérationnels ou en instance d'une ontologie médicale (telle que Asbru). Cette approche permet donc de formaliser le Guide de Bonnes Pratiques mais n'est pas un outil pour le développement complet d'un système

⁴⁵ <http://www.w3.org/TR/xslt> (dernier accès le 10 juin 2006).

⁴⁶ http://www.openclinical.org/gmm_stepper.html (dernier accès le 10 juin 2006).

d'aide à la décision (Svatek, Kroupa & Ruzicka, 2001). Stepper permet de visualiser l'interconnexion entre le texte source et le texte formalisé, mais le balisage reste un processus « manuel » et consiste à délimiter des blocs initiaux de « connaissances ».

2.7.3 GMT (Guideline Markup Tool) / DELT/A

GMT pour *Guideline Markup Tool* (Kosara et al., 2002; Moser, 2004; Votruba et al., 2004), appelé plus récemment DELT/A, est associé au projet PROTOcure⁴⁷. Cette interface a pour vocation de faciliter le passage du texte des Guides de Bonnes Pratiques vers une représentation formelle basée sur XML.

Une des fonctionnalités principales consiste à faire le lien entre le texte source et sa représentation formelle. Cette interface permet également d'éditer des protocoles Asbru. GMT a pour particularité la possibilité de faire des connections entre le texte original du Guide de Bonnes Pratiques et sa version informatisée. Par exemple, durant la phase de validation du Guide de Bonnes Pratiques représenté formellement, si la représentation formelle conduit à des contradictions, une fonctionnalité utile consisterait à revenir au texte original. GMT intègre cette fonctionnalité en créant des liens entre le Guide de Bonnes Pratiques original et sa représentation formelle, plus précisément dans le modèle Asbru. On peut donc connaître l'origine d'une valeur spécifique du fichier XML Asbru puisque le programme intégré à GMT permet de rechercher la valeur correspondante dans le fichier HTML contenant le Guide de Bonnes Pratiques original.

Une deuxième fonctionnalité concerne l'utilisation de macros, caractérisées par la combinaison de plusieurs éléments XML (les éléments Asbru) qui sont généralement utilisés ensemble, permettant la construction d'algorithmes souvent décrits dans les Guides de Bonnes Pratiques.

GMT supporte donc les tâches suivantes :

- (i) Écriture et amélioration des Guides de Bonnes Pratiques : une version Asbru du Guide de Bonnes Pratiques est créée, et des liens sont ajoutés entre les parties correspondantes du Guide de Bonnes Pratiques et le fichier Asbru créé.
- (ii) Compréhension du Guide de Bonnes Pratiques Asbru : identification de la provenance des valeurs du fichier Asbru dans le Guide de Bonnes Pratiques, ainsi que les parties du Guide de Bonnes Pratiques qui ont été utilisées.
- (iii) Structuration Asbru : liste d'éléments Asbru, plus précisément les macros qui ont besoin d'être structurées et qui permettent l'écriture des plans.

L'interface GMT est séparée en trois parties. Le texte original du Guide de Bonnes Pratiques en HTML est affiché en haut à gauche de la Figure 2.27, le fichier XML représentant le fichier XML Asbru est affiché en haut à droite, et la dernière partie en bas de l'interface est réservée aux macros (à gauche la macro-structure et à droite la pré-visualisation de la macro sélectionnée).

⁴⁷ <http://www.keg.uji.es/protopub> (dernier accès le 10 juin 2006).

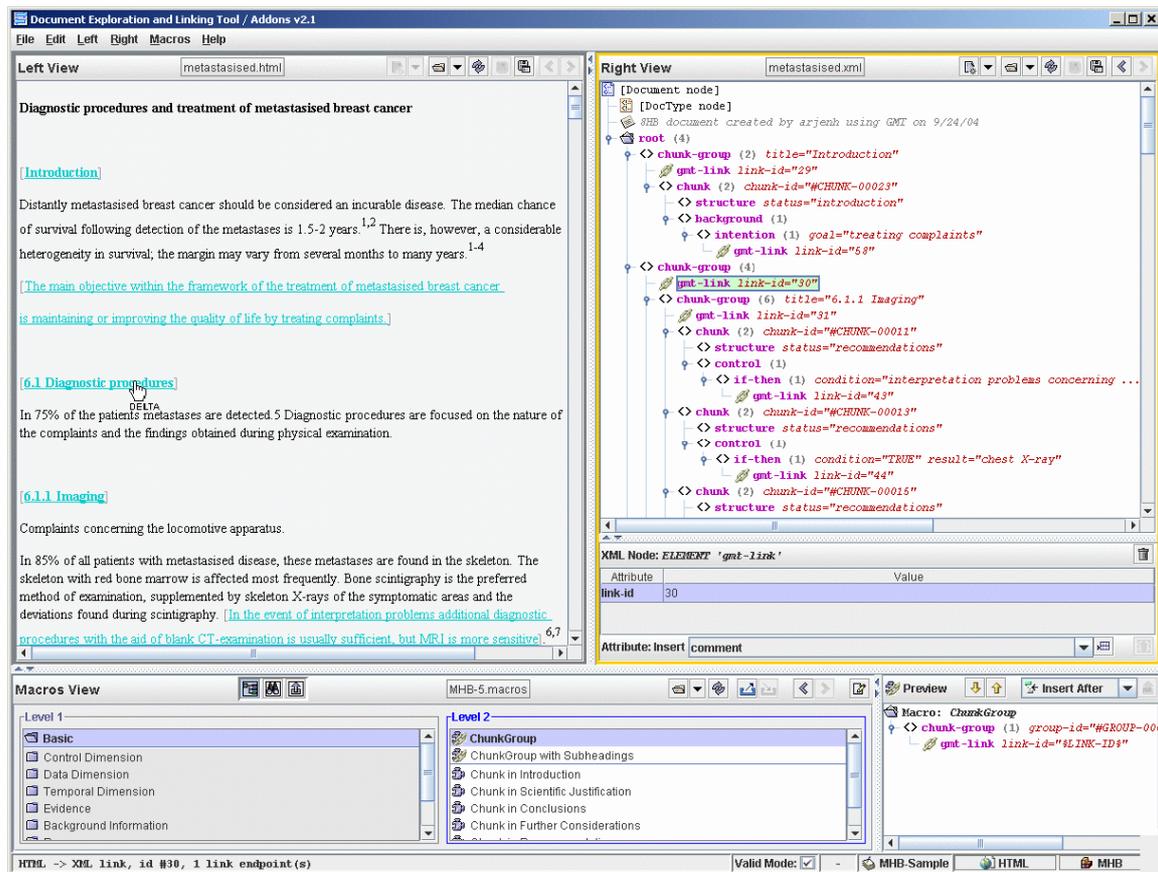


Figure 2.27 – Aperçu de l'interface GMT représentant l'encodage d'un Guide de Bonnes Pratiques en Asbru⁴⁸.

2.7.4 DeGeL / URUZ

Le *framework* DeGeL (*Digital Electronic Guideline Library*)⁴⁹ (Shahar et al., 2003) intègre un ensemble d'outils pour supporter le développement et l'implémentation des Guides de Bonnes Pratiques. Par exemple, lors de la spécification des Guides de Bonnes Pratiques, URUZ (et GESHER⁵⁰) permet de vérifier le processus de spécification (syntaxe) et la validation du Guide de Bonnes Pratiques en fonction de ses objectifs (sémantique).

Lors de l'exécution, VisiGuide, Vaidurya et IndexiGuide permettent de déterminer l'éligibilité d'un patient et l'applicabilité d'un Guide de Bonnes Pratiques, et de visualiser un ou plusieurs Guides de Bonnes Pratiques potentiellement applicables au profil du patient. QualiGuide permet de vérifier la qualité de l'évaluation du Guide de Bonnes Pratiques formalisé, et Spock permet de le modifier et d'évaluer son efficacité. URUZ (ou GESHER dans sa version utilisable en ligne) est un environnement graphique (Figure 2.28) qui permet de structurer le contenu des Guides de Bonnes Pratiques et de le transformer vers un format exécutable tel que Asbru.

⁴⁸ <http://ieg.ifs.tuwien.ac.at/projects/delta/> (dernier accès le 10 juin 2006).

⁴⁹ http://www.openclinical.org/gmm_degel.html (dernier accès le 10 juin 2006).

⁵⁰ GESHER est l'éditeur web d'URUZ.

Chaque élément d'Asbru a une représentation graphique spécifique, incluant ses caractéristiques (par exemple, le corps du plan (*plan body*)). URUZ supporte donc le balisage du texte pour en faire un texte structuré en utilisant des balises sémantiques dans le but de cibler le langage de représentation (par exemple, Asbru ou GEM). URUZ comporte une vue pour chaque niveau de structuration, et l'arborescence est fonction des connaissances « pertinentes » du niveau dans lequel on se situe. Par exemple, l'expert fait glisser l'élément sélectionné « *filter-precondition* » dans la zone d'édition.

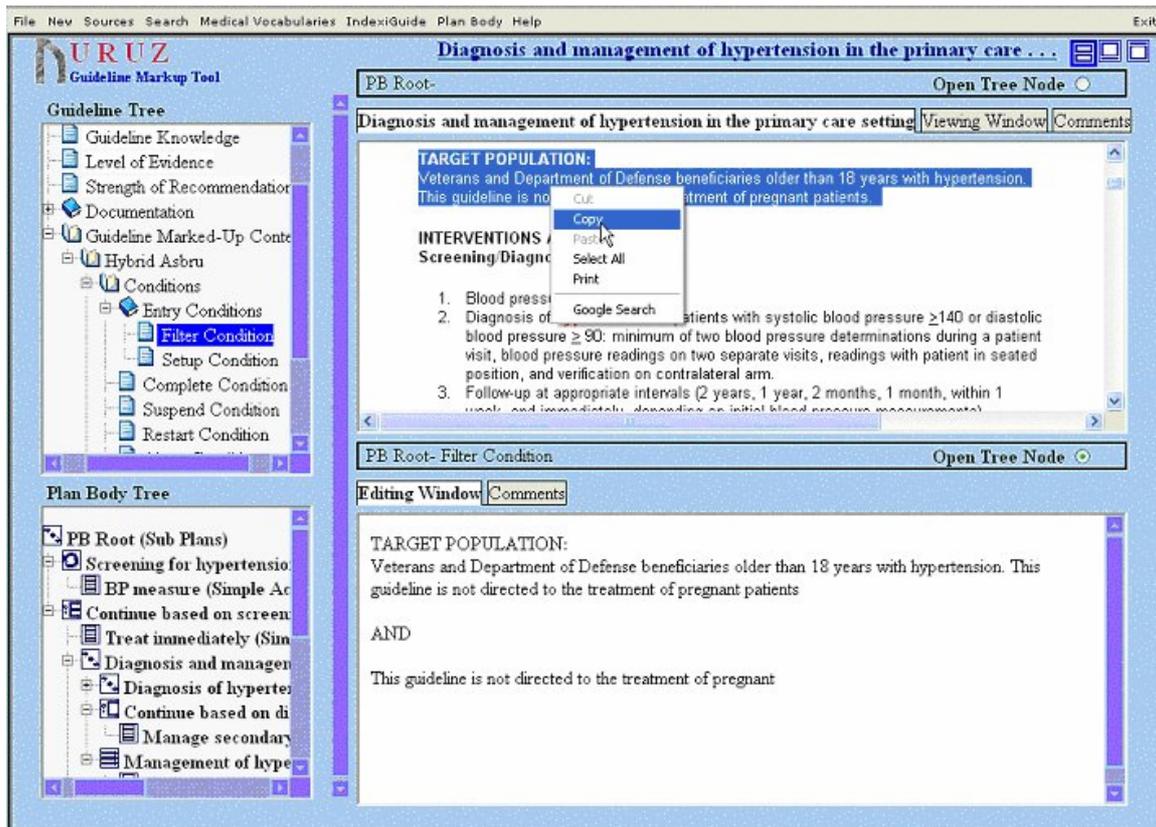


Figure 2.28 – Interface d'URUZ⁴⁹ permettant le passage du texte vers le langage Asbru et générant automatiquement le fichier XML correspondant.

2.8 Conclusion

Nous avons présenté dans ce chapitre les principales approches de l'informatisation des Guides de Bonnes Pratiques. Différents formalismes pour représenter le contenu de ces documents ont été introduits, en insistant sur leur diversité et sur la difficulté d'encoder de tels documents dans des formalismes décrivant les processus cliniques. Néanmoins, en amont de ces formalismes, les approches documentaires se sont intéressées au document du point de vue de leur structure pour préserver toutes les informations textuelles contenues dans le document lors du passage des connaissances vers des formalismes de représentation des connaissances. Ces approches se focalisent sur l'ajout d'informations au texte en utilisant le principe de balisage. GEM est le modèle le plus connu et a été reconnu

2.8. Conclusion

comme un standard pour structurer les Guides de Bonnes Pratiques. Néanmoins, le passage du texte vers ce modèle, appelé encodage manuel, comporte également des limites telles que la variation du document créé dépendant de la personne en charge de l'encodage, la perte de temps, et les problèmes d'interprétation dus au fait que le texte est mal structuré et est sujet à discussion quant au contenu.

Les interfaces graphiques actuelles facilitent le passage du texte vers un format informatique mais n'agissent pas sur le contenu du document. Ainsi, les experts n'ont pas besoin d'apprendre un langage informatique pour pouvoir représenter le contenu des Guides de Bonnes Pratiques. Le Tableau 2.5 présente une synthèse des caractéristiques des interfaces présentées précédemment.

Un aspect important de l'informatisation des Guides de Bonnes Pratiques consiste à assister les médecins dans la production des Guides de Bonnes Pratiques et dans leur encodage correct dans des formats tels que GEM. Nous avons au cours de nos travaux observé de nombreuses limites durant l'encodage manuel des Guides de Bonnes Pratiques (Georg, Séroussi & Bouaud, 2003). Des outils pour faciliter le passage du texte vers un format informatique ont été développés mais reposent actuellement sur un processus « manuel », fournissant uniquement une interface graphique à l'utilisateur.

C'est dans ce contexte que nous proposons l'environnement G-DEE (*Guideline Document Engineering Environment*) dédié à l'étude des Guides de Bonnes Pratiques et intégrant des fonctions de traitement automatique de texte telles que l'identification de marqueurs linguistiques pour reconnaître automatiquement les recommandations dans le document (Georg & Jaulent, 2005). Le balisage automatique de tels marqueurs facilite autant la structuration du document, que son élaboration, sa formalisation et son intégration dans des systèmes d'aide à la décision. L'originalité de cette approche tient également dans le fait que les techniques d'extraction de contenu ont la plupart du temps été pensées en termes d'extraction de variables cliniques et en termes d'ontologies. Nous proposons, comme développé par la suite, de les utiliser dans un premier temps pour la reconnaissance d'éléments structurants.

Tableau 2.5 – Caractéristiques des interfaces dédiées aux approches documentaires.

Interface	Texte source visible	Texte source modifié	Balises manuelles et dépendantes de l'expert	Manipulation du texte	Étapes pour structurer le document	Document généré après structuration
GEM-Cutter	oui	non	oui	sélectionner / glisser la portion de texte dans des catégories	une seule étape	fichier XML représentant le texte encodé en GEM
Guide-X / Stepper	oui	non	oui	sélectionner / glisser la portion de texte dans des catégories	5 étapes	fichier XML représentant le texte encodé en Asbru
GMT / DELT/A	oui	non	oui	sélectionner / glisser la portion de texte dans des catégories / lien dynamique avec le texte source	une seule étape	fichier XML représentant le texte encodé en Asbru
DeGeL / Uruz / GESHER	oui	non	oui	sélectionner / glisser la portion de texte dans des catégories	une seule étape	format propriétaire Asbru

Chapitre 3

Caractérisation linguistique des Guides de Bonnes Pratiques

Dans ce chapitre nous cherchons à caractériser les phénomènes linguistiques permettant de structurer les Guides de Bonnes Pratiques, et plus particulièrement permettant de reconnaître automatiquement les recommandations contenues dans les Guides de Bonnes Pratiques. Nous introduisons la notion d'opérateurs déontiques dont la forme linguistique (verbes déontiques de type recommander, devoir, envisager) constitue l'expression des recommandations. Pour ce faire, l'étude des champs lexicaux des verbes déontiques ainsi que des analyses distributionnelles des opérateurs déontiques dans les textes seront présentées.

3.1 Opérateurs déontiques dans les Guides de Bonnes Pratiques

3.1.1 Verbes déontiques dans la littérature

Les Guides de Bonnes Pratiques cliniques appartiennent à une catégorie de textes très étudiée qu'on désigne généralement sous le terme de textes réglementaires. Ces textes se caractérisent par la nature de l'information contenue mais également par la présence de phénomènes linguistiques caractéristiques. En effet, les textes réglementaires expriment par nature des modalités comme l'obligation, la permission ou l'interdiction, qu'on appelle modalités déontiques. Ces textes réglementaires sont caractérisés par la fréquence particulière des verbes déontiques tels que devoir, pouvoir et interdire (Kalinowski, 1996).

D'après Kalinowski, la norme-énoncé désigne un état de choses de nature relationnelle représentant une relation existante entre un homme (sujet d'action) et une action. Par exemple, « une action qu'il doit faire, qu'il ne doit pas faire ou, plus précisément doit ne pas faire, qu'il a le droit de faire, qu'il n'a pas le droit de faire ou qu'il peut (au sens d'avoir la liberté de choix) faire et ne pas faire ». La théorie des normes correspond à des expressions du langage naturel servant, selon les règles linguistiques, à énoncer les normes-jugements. Ces normes peuvent être exprimées de plusieurs manières, par exemple, à l'aide de termes normatifs, tels que « il faut ... », « il doit ... », « ... est obligé ... ».

La recommandation est une forme linguistique qui correspond à une injonction d'adopter une certaine conduite en fonction d'une situation clinique et de connaissances médicales sous-jacentes. Considérons quelques exemples de recommandations.

Exemple 1 : *En cas de traitement de l'hypertension artérielle, il est recommandé d'utiliser préférentiellement la perfusion intraveineuse (IV) pour un ajustement tensionnel précis. Les voies intramusculaire et sublinguale sont à éviter. L'utilisation préférentielle de l'urapidil ou du labétalol ou de la nicardipine est recommandée, en évitant les doses de charge (accord professionnel)*⁵¹.

Exemple 2 : *Les troubles de la déglutition doivent être recherchés systématiquement avant la première alimentation, car ils exposent au risque de pneumopathie d'inhalation (grade B)*⁵¹.

Exemple 3 : *Chez les patients dont le traitement peut induire des modifications de la créatinine (diurétiques, IEC, ou ARA II), un dosage annuel de la créatinine et du ionogramme sanguin est recommandé*⁵².

⁵¹ Extrait du Guide de Bonnes Pratiques « Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral – Aspects médicaux » ; ANAES, Septembre 2002. (<http://www.has-sante.fr> (dernier accès le 10 juin 2006)).

⁵² Extrait du Guide de Bonnes Pratiques « Prise en charge des patients adultes atteints d'hypertension artérielle essentielle » ; ANAES, Avril 2000. (<http://www.has-sante.fr> (dernier accès le 10 juin 2006)).

Moulin et Rousseau ont proposé, dans le projet A.C.A.T (Acquisition des Connaissances et Analyse des Textes), une approche d'extraction de connaissances à partir de textes légaux basée sur les verbes déontiques qui sont utilisés pour structurer le contenu de textes prescriptifs (Moulin & Rousseau, 1990). Moulin et Rousseau ont montré que le contenu d'énoncés prescriptifs est spécifié par des propositions normatives, qui se manifestent, dans les textes qu'ils ont étudiés, par des verbes tels que pouvoir, devoir et interdire, en d'autres termes des verbes déontiques. Ils ont montré que l'extraction de connaissances à partir de textes légaux basée sur les opérateurs déontiques est un moyen efficace de résoudre les problèmes d'acquisition de connaissances à partir de textes, sans avoir à prendre en compte le niveau de détail des énoncés, mais en les utilisant pour structurer le document. Ces résultats vont servir de point de départ à notre approche et nous nous proposons d'étudier le rôle des opérateurs déontiques dans les Guides de Bonnes Pratiques cliniques et par tant l'utilisation qu'il serait possible d'en faire dans le cadre de traitements automatiques.

On peut également noter, sans pour autant en tirer de conclusions quant au rapport avec les Guides de Bonnes Pratiques à l'énonciation orale, qu'il existe une certaine proximité entre la recommandation et un acte de langage (Austin, 1962; Searle, 1965) et plus particulièrement les actes définis par Austin comme des « exercitifs » qui consistent à formuler une décision en faveur ou à l'encontre d'une suite d'actions. Par exemple : ordonner, commander, plaider pour, supplier, recommander, implorer, conseiller, et les verbes tels que nommer, déclarer une séance ouverte, avertir, proclamer. Un rapprochement entre performatifs et déontiques a également été proposé plus récemment par Kerbrat-Orecchioni⁵³.

3.1.2 Méthodologie pour l'identification des opérateurs déontiques

Les opérateurs déontiques se traduisent linguistiquement par la présence de locutions verbales organisées autour de verbes déontiques. En dehors des verbes déontiques canoniques exprimant la permission, l'interdiction et l'obligation (comme recommander, devoir, et interdire), il importe de pouvoir reconnaître en contexte d'autres expressions déontiques. Nous avons, dans un premier temps, établi que 100% des occurrences de recommander correspondent bien à une expression déontique signalant une recommandation élémentaire (Tableau 3.6). Cette expression peut donc constituer notre référence pour évaluer le rôle déontique d'autres expressions. Pour ce faire, nous allons comparer les environnements lexicaux des différents opérateurs candidats à l'environnement lexical de l'opérateur de référence *recommander*. Pour ce faire nous utilisons TropesTM⁵⁴ sur un corpus constitué de 47 Guides de Bonnes Pratiques (cf. Annexe 2) de l'HAS⁵⁵ contenant 616 verbes. Les co-occurrences du verbe *recommander* sont représentées par la Figure 3.30. Nous pouvons observer que les contextes gauche ou droit font

⁵³ <http://www.reds.msh-paris.fr/communication/textes/dcad.htm> (dernier accès le 10 juin 2006).

⁵⁴ <http://www.acetic.fr> (dernier accès le 10 juin 2006).

⁵⁵ <http://www.has-sante.fr> (dernier accès le 10 juin 2006).

intervenir des acteurs tels que patient, ainsi que des traitements et des grades. L'occurrence d'un même champ lexical à gauche et à droite s'explique par l'alternance de voix passive et active dans les textes.

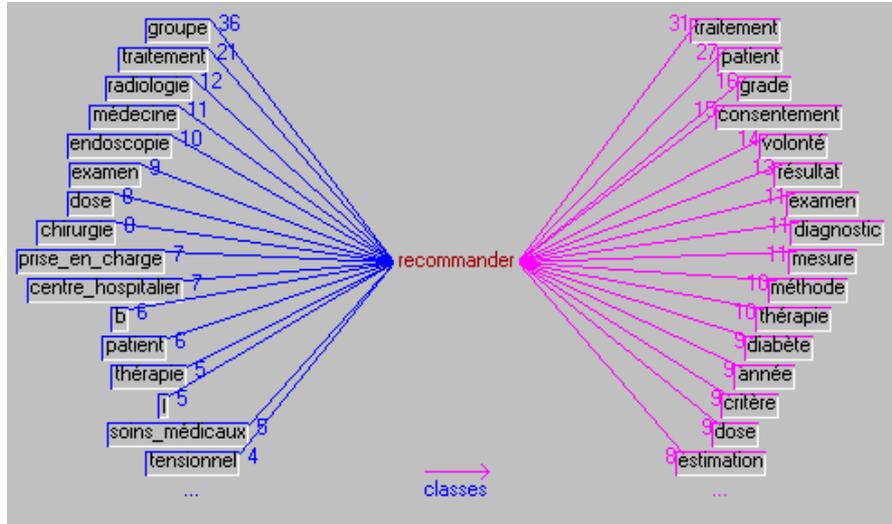


Figure 3.29 – Co-occurrences du verbe recommander dans le corpus.

Nous procédons de même avec les co-occurrences du verbe devoir, représentées par la Figure 3.30, pour lesquelles les champs lexicaux sont similaires à ceux du verbe recommander. Par exemple, les acteurs tels que patient, centre hospitalier, médecin ; la thérapeutique telle que le traitement, la prise en charge ; le grade (caractéristique d'une recommandation élémentaire) ; les éléments diagnostiques tels que les examens.

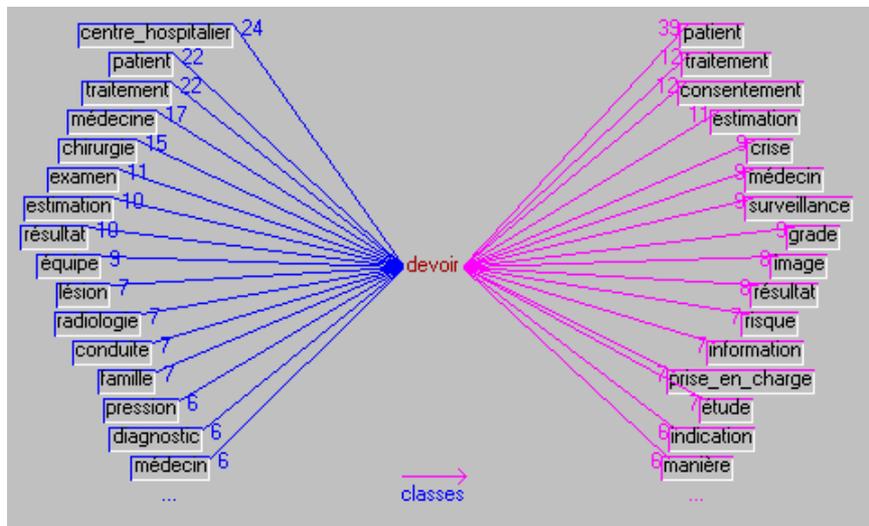


Figure 3.30 – Co-occurrences du verbe devoir.

3.1. Opérateurs déontiques dans les GBP

De manière qualitative, on observe une concordance de 77% des champs lexicaux, plus précisément les champs similaires à ceux du verbe recommander. De manière quantitative, on décompte 56 occurrences de la catégorie thérapeutique, 76 occurrences de la catégorie acteur, 42 occurrences pour les éléments diagnostiques, et 9 occurrences pour le grade. On observe une similarité par rapport au nombre d'occurrences de ces catégories pour le verbe recommander (104 pour la thérapeutique, 76 pour les acteurs, 73 pour les éléments diagnostiques et 16 pour le grade).

Nous pouvons mettre ces résultats en relation avec le pourcentage pour lequel le verbe devoir indique une recommandation et qui correspond à 89% de ces occurrences (779 occurrences dans le corpus cf. Tableau 3.6). On peut également noter que la construction passive du verbe devoir concerne 411 occurrences (52,75%), et la voix active 368 occurrences (47,25%).

On considère donc ce verbe comme un verbe déontique du fait de sa proximité au niveau des champs lexicaux avec le verbe recommander ainsi que sa forte représentativité dans le corpus.

Nous avons aussi recherché les similitudes des co-occurrences du verbe *pouvoir* avec le verbe *recommander* (Figure 3.31). On observe également que les champs lexicaux de ce verbe concernent les acteurs (tels que le *patient*), la thérapeutique (comme *traitement*), l'association (de classes thérapeutiques), et des éléments diagnostiques tels que la radiologie. On observe une concordance de 65% des champs lexicaux par rapport à ceux du verbe *recommander*. De manière quantitative, on décompte 80 occurrences pour la catégorie thérapeutique, 45 pour celle des acteurs, 42 pour les éléments diagnostiques et 7 pour le grade. On peut également observer une similarité du nombre d'occurrences de ces catégories par rapport au nombre d'occurrences des catégories du verbe recommander, par exemple la catégorie thérapeutique (80 occurrences pour pouvoir et 104 pour recommander).

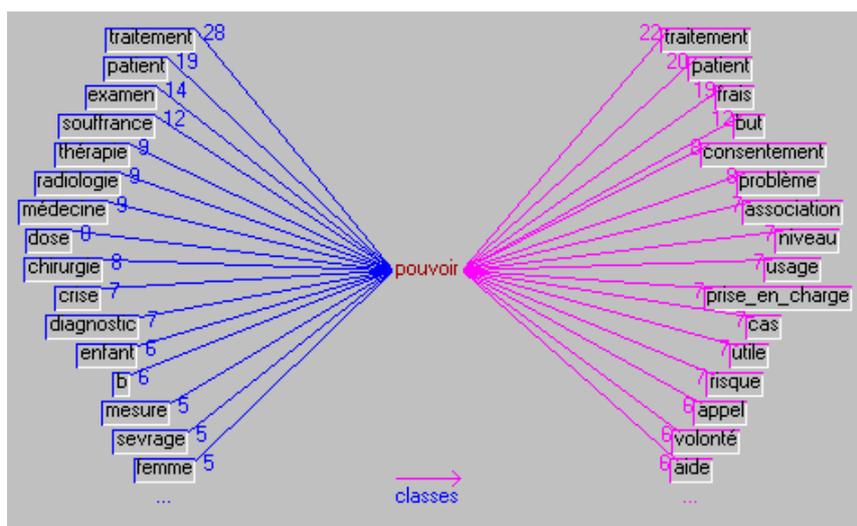


Figure 3.31 - Co-occurrences du verbe pouvoir.

De la même manière que précédemment, nous pouvons mettre ces résultats en relation avec le pourcentage pour lequel le verbe pouvoir indique une recommandation

correspondant à 76% de ces occurrences (718 occurrences dans le corpus cf. Tableau 3.6). La construction passive du verbe pouvoir concerne 424 occurrences (59%), et la voix active 294 occurrences (41%). On considère donc le verbe pouvoir comme un verbe déontique du fait de sa proximité au niveau des champs lexicaux avec le verbe recommander.

D'autres verbes ont été étudiés de cette manière, et l'on a pu établir une liste d'opérateurs déontiques à partir de ceux dont le contexte se rapproche le plus du verbe déontique « canonique » *recommander* : devoir, pouvoir, nécessiter, indiquer, justifier, conseiller, préconiser, imposer, choisir, bénéficier, privilégier, préférer, insister, effectuer, envisager, éviter, déconseiller, exister. L'analyse des contextes peut aussi révéler des cas particuliers. Par exemple, le verbe débiter (Figure 3.32) présente un contexte très semblable à celui d'un opérateur déontique. Pour les 32 occurrences de ce verbe, on reconnaît une association directe à un verbe déontique, comme dans les occurrences suivantes : « peut débiter » ou « il est recommandé de débiter ».

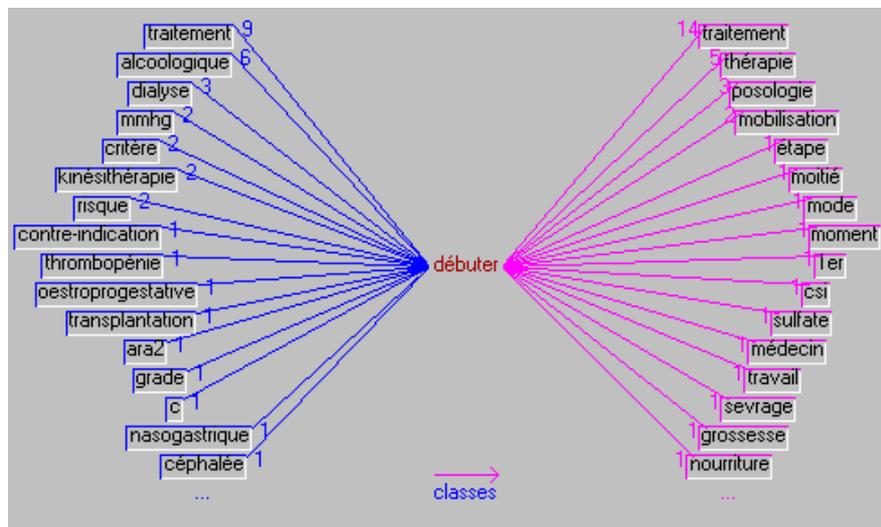


Figure 3.32 – Co-occurrences du verbe débiter.

Par la méthode des champs lexicaux similaires au verbe recommander, on a pu éliminer les verbes qui ne se comportaient pas dans le corpus comme des verbes déontiques : considérer, démontrer, autoriser, adopter, encourager, exclure, veiller, vérifier, faire.

Ci-dessous le contexte lexical d'un verbe non déontique, *veiller* qui se différencie nettement de celui du verbe *recommander*.

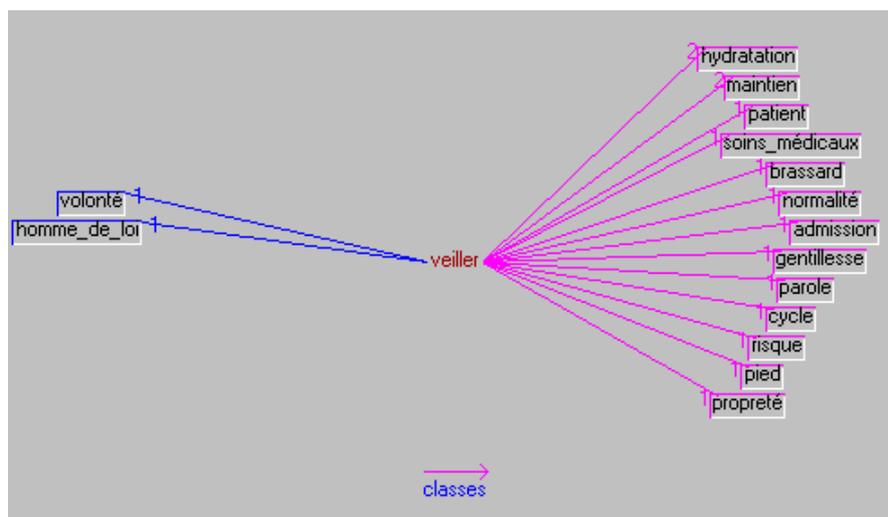


Figure 3.33 – Co-occurrences du verbe veiller.

Le cas du verbe *falloir* est également spécifique (Figure 3.34). On observe (par une analyse manuelle du corpus) que ce verbe indique pour 66% de ces occurrences des recommandations (Tableau 3.6).

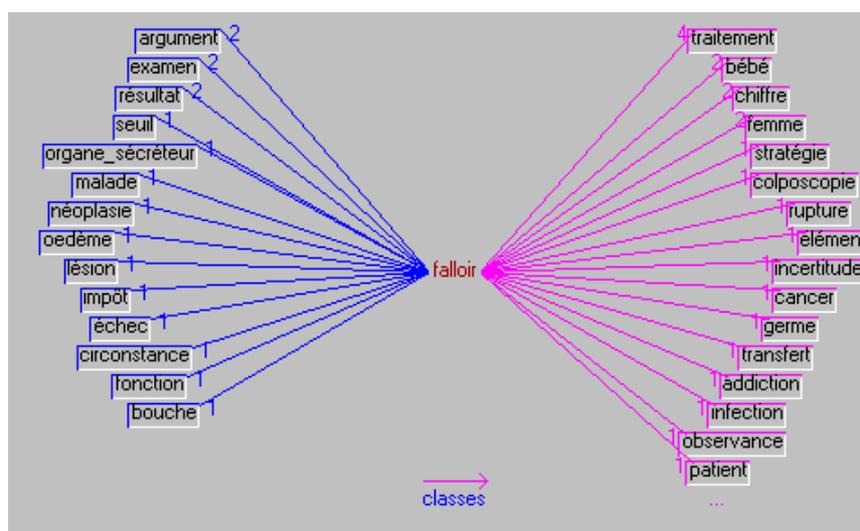


Figure 3.34 – Co-occurrences du verbe falloir.

Néanmoins, ses champs lexicaux ne présentent que 34,5% de similarités avec ceux du verbe recommander. L'explication doit se trouver dans le fait que falloir se comporte comme un verbe support d'autres verbes déontiques. Par exemple, « Il faut éviter toute hyperhydratation, notamment intracellulaire. » ou « Il faut rechercher une addiction à un autre médicament (en particulier les benzodiazépines), au café, au tabac, à l'alcool, aux drogues. »

Tableau 3.6 – Occurrences des verbes déontiques et celles pour lesquelles les verbes indiquent des recommandations élémentaires.

Verbe déontique	Nombre d'occurrences indiquant une recommandation	Nombre d'occurrences total	% d'occurrences du verbe indiquant une recommandation
Recommander	897	897	100%
Devoir	779	874	89%
Pouvoir	718	946	76%
Proposer	200	392	51%
Envisager	59	141	42%
Indiquer	58	93	62%
Falloir	31	47	66%
Bénéficier	29	40	72,5%
Conseiller	29	35	83%
Eviter	27	129	21%
Exister	25	137	18%
Imposer	22	30	73%
Préconiser	21	36	58%
Privilégier	18	32	56%
Préférer	12	21	57%
Déconseiller	10	10	100%
Insister	10	11	91%

3.2 Distributions des opérateurs déontiques dans les textes

L'analyse des phénomènes linguistiques ne se limite pas au palier de la phrase ; la reconnaissance de certains phénomènes linguistiques au niveau textuel peut être particulièrement utile à la formulation de mécanismes de traitement automatique. Ce type d'analyse est cependant rendu difficile par la complexité des connaissances linguistiques textuelles (Rastier, 1989). Dans une perspective plus modeste, nous avons recherché des régularités de structure des Guides de Bonnes Pratiques en utilisant un outil d'analyse simple, c'est-à-dire la distribution des opérateurs déontiques (Georg, Colombet & Jaulent, 2005). Afin d'éviter d'introduire des biais dans cette analyse, nous avons étudié un seul opérateur déontique « canonique », correspondant au verbe recommander, pour lequel nous avons établi plus haut que toutes ses occurrences correspondaient bien à des opérateurs déontiques.

3.2. Distributions des opérateurs déontiques

Nous avons étudié (au moyen du logiciel TropesTM)⁵⁶ la distribution suivant le déroulement linéaire du texte du verbe déontique *recommander*, pour dix Guides de Bonnes Pratiques⁵⁷ produits par des auteurs (groupes de travail) différents et sur une période de plusieurs années (1998-2004). Les résultats sont présentés sous la forme de diagrammes de fréquence (Figure 3.35) avec, en abscisse, les segments du texte numérotés du début vers la fin du document (la taille moyenne de ces segments, déterminés par le logiciel TropesTM, représente environ 150 mots) et, en ordonnée, la fréquence (nombre d'occurrences) du verbe déontique *recommander*. Nous nous intéressons exclusivement à la forme de la distribution, dans le but de rechercher des régularités naturelles de structure qui caractériseraient le « genre textuel » des Guides de Bonnes Pratiques.

La Figure 3.35 montre une grande diversité dans la forme des distributions de l'opérateur déontique à travers les documents. Ces distributions confirment la fréquence significative de l'opérateur déontique mais ne permettent de conclure à aucune régularité de structure du texte lui-même, tant les formes de ces distributions sont variées.

Nous sommes donc amenés à faire l'hypothèse qu'un Guide de Bonnes Pratiques est avant tout une juxtaposition de recommandations. Il nous reste à pratiquer une expérience complémentaire afin de vérifier que les différences entre distributions ne sont pas le résultat de variations de vocabulaire dans les opérateurs déontiques. Une constatation simple peut nous aider à résoudre ce point, qui est la corrélation observée entre les occurrences de différents verbes déontiques. Bien que nous ne détaillions pas ici cette expérience, les résultats de la Figure 3.36 en contiennent l'essentiel : celle-ci compare les distributions des verbes déontiques *recommander*, *devoir* et *pouvoir* dans quatre des textes étudiés plus haut (Figure 3.35). Elle montre que les verbes déontiques (et donc les recommandations) se concentrent dans les mêmes sections des documents : les différences de distribution observées ne peuvent donc se résumer à des variantes lexicales mais dénotent bien une grande variabilité de structure des Guides de Bonnes Pratiques.

⁵⁶ <http://www.acetic.fr> (dernier accès le 10 juin 2006).

⁵⁷ <http://www.has-sante.fr> (dernier accès le 10 juin 2006).

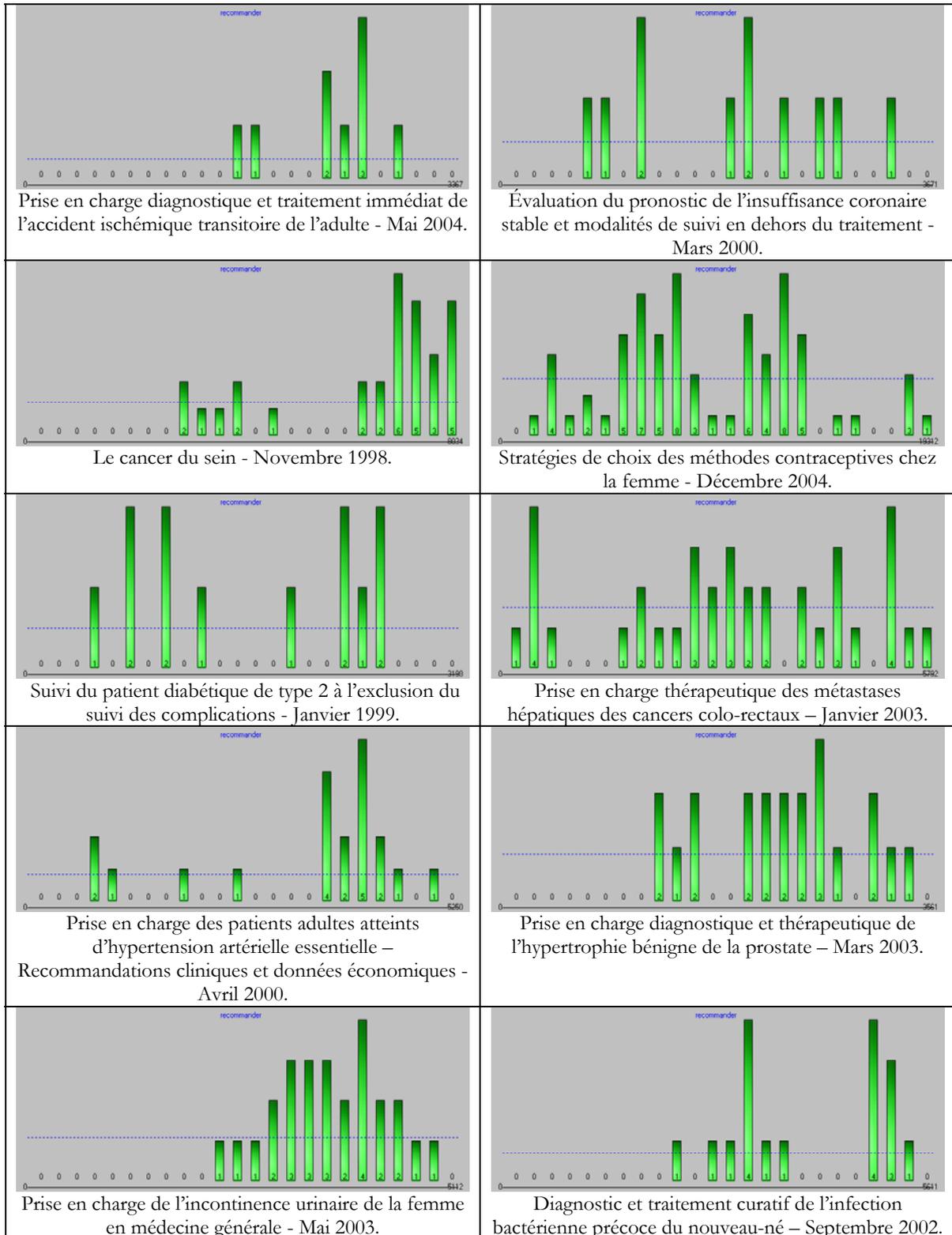


Figure 3.35 – Distribution du verbe *recommander* dans dix Guides de Bonnes Pratiques de l'ANAES⁵⁷.

3.2. Distributions des opérateurs déontiques

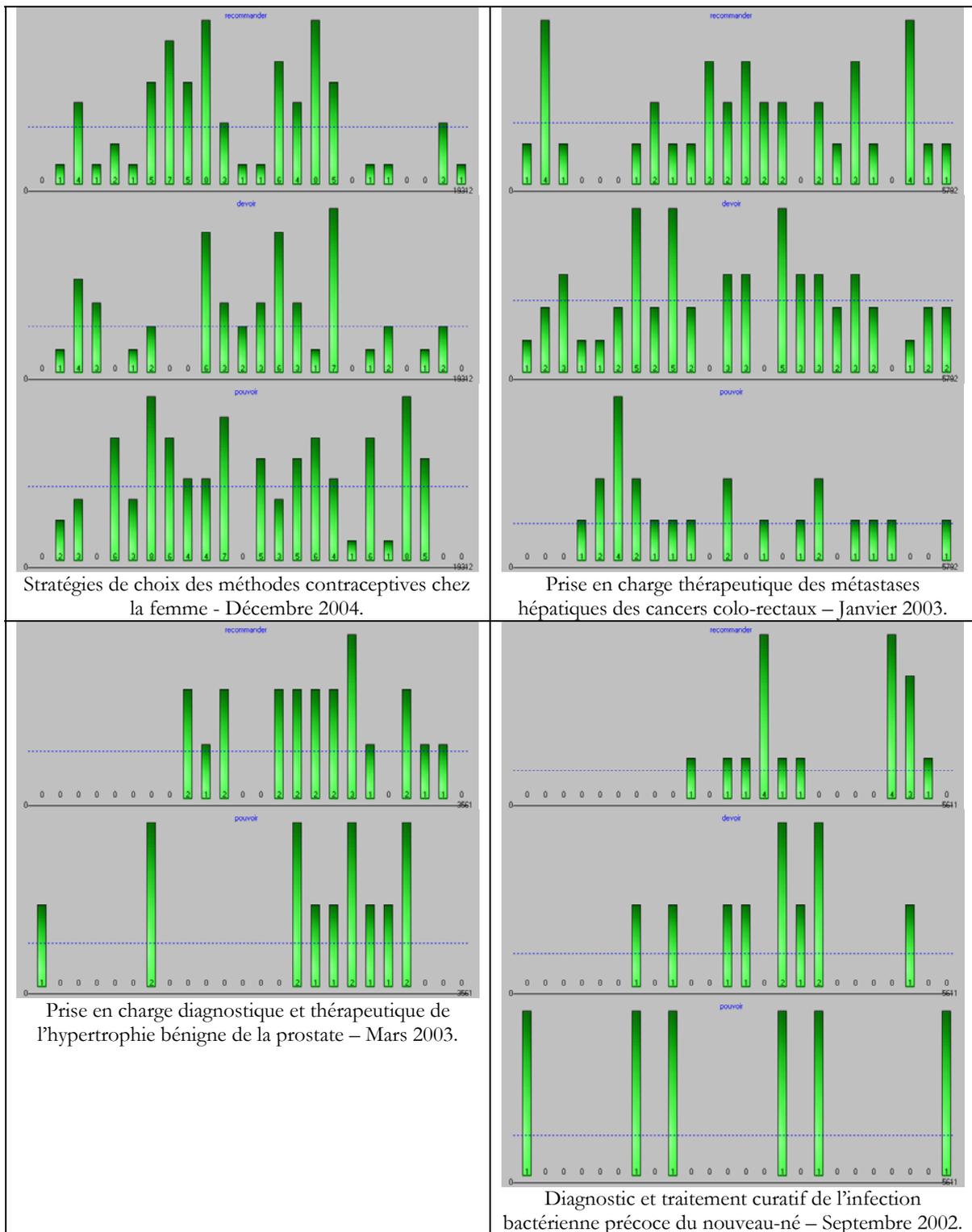


Figure 3.36 - Distribution comparative des verbes déontiques principaux (*recommander*, *devoir*, *pouvoir*) dans quatre Guides de Bonnes Pratiques de l'ANAES⁵⁷.

3.3 Conclusion

Nous avons montré que les Guides de Bonnes Pratiques, comme d'autres textes réglementaires sont caractérisés par la fréquence d'expressions linguistiques particulières, les opérateurs déontiques.

Nous avons également montré l'équivalence qui existe entre opérateurs déontiques et recommandations élémentaires ; en d'autres termes, les unités de connaissances fondamentales véhiculées par les textes correspondent aux structures linguistiques que sont les verbes déontiques.

Enfin, une étude de la distribution des recommandations dans les textes ne suggère pas de régularité de structure dans les Guides de Bonnes Pratiques ; nous devons abandonner l'idée d'utiliser des régularités textuelles pour guider les traitements automatiques. Nous ferons donc, pour la suite de notre étude, l'hypothèse heuristique que les Guides de Bonnes Pratiques se comportent avant tout comme un assemblage de recommandations individuelles. La reconnaissance de ces recommandations peut donc servir de base à une structuration des Guides de Bonnes Pratiques, pour peu qu'on puisse disposer de techniques d'analyse linguistique automatique identifiant les opérateurs déontiques leur correspondant, c'est-à-dire des verbes déontiques opérant sur des portions de texte.

Chapitre 4

Approches linguistiques

Ce chapitre est dédié à la description de notre approche technique, consistant à inclure des techniques empruntées à l'Extraction d'Informations à une approche documentaire plus classique comme principe servant à la structuration automatique.

4.1 Caractérisation de l'approche

L'analyse linguistique exposée dans le chapitre précédent a mis en évidence les concepts qui permettent de caractériser et d'analyser les textes des Guides de Bonnes Pratiques. La description linguistique, au-delà des phénomènes qu'elle identifie dans les textes, peut aussi servir de base à une analyse informatique des documents. C'est en effet elle qui permet la formalisation (par exemple sous forme de motifs syntaxiques ou syntaxico-sémantiques) indispensable à une reconnaissance automatique d'éléments significatifs du texte, comme les occurrences de recommandations.

Dans ce chapitre, nous nous attachons à caractériser notre approche en termes techniques⁵⁸. Les problèmes abordés concernent principalement la structuration des textes à travers la reconnaissance de certains contenus, et les techniques informatiques que nous pouvons convoquer pour cette analyse automatique.

Nous présentons d'abord les travaux réalisés dans le domaine du traitement du document et qui s'intéressent principalement à la mise en forme du texte (comme la mise en relief de certaines sections). Ce domaine s'est développé de façon significative ces dernières années avec les langages à base de balises qui offrent une approche générique et flexible à travers les différentes techniques de balisage et les transformations qu'elles permettent. Nous nous intéresserons ensuite à une forme particulière de traitement du contenu linguistique des textes, l'Extraction d'Informations. Cette approche de l'analyse automatique du langage naturel permet de reconnaître de manière automatique mais sélective certains concepts quelle que soit la forme linguistique de leur occurrence dans le texte.

Le schéma ci-dessous représente le point de vue général de notre approche, s'inspirant à la fois des techniques de traitement du document (par la modification du document source basée sur une interprétation externe) et des techniques d'Extraction d'Informations (pour l'identification de recommandations qui servira de point de départ à la structuration du texte). Nous proposons de combiner ces deux techniques, ce qui conférera un caractère « intelligent » (c'est-à-dire basée sur une détection automatique du contenu linguistique) à la structuration du document. Du point de vue de la stratégie d'analyse, le texte source sera à la fois le point de départ et le point d'arrivée de l'interprétation automatique ; ceci contraste avec les techniques d'analyse ou d'Extraction d'Informations qui visent à instancier des représentations informatiques à partir du texte.

⁵⁸ Ce chapitre ne prétend pas constituer un historique de toutes les techniques de traitement du document et d'Extraction d'Informations, ni même un état de l'art exhaustif. Il sélectionne les techniques les plus représentatives dont nous nous sommes inspirés pour la partie informatique de nos travaux.

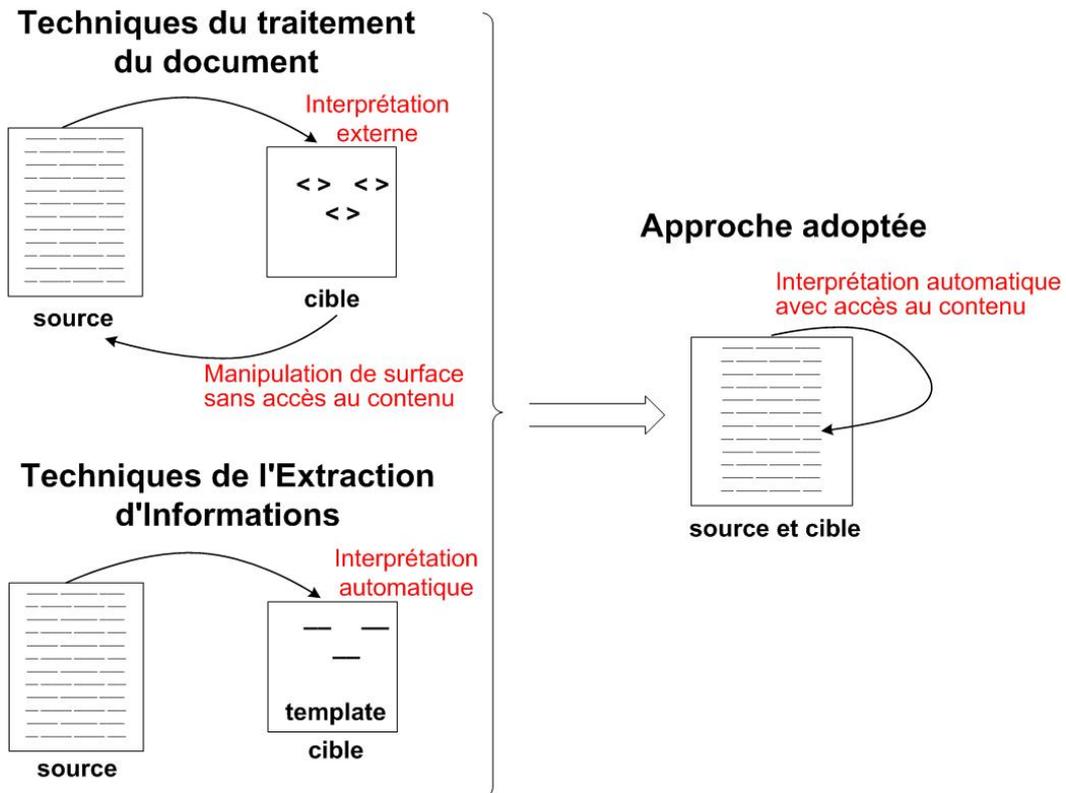


Figure 4.37 – Représentation générale de l'approche adoptée qui correspond à la fusion des techniques issues du traitement du document et de l'Extraction d'Informations.

4.2 Les techniques du traitement du document

Les techniques de traitement du document appelées aussi *document engineering* s'intéressent plus particulièrement à la structure du document et non à son contenu, et comprennent l'ensemble des méthodes supportant la structuration, la création et la reconfiguration de documents numériques multimédia⁵⁹.

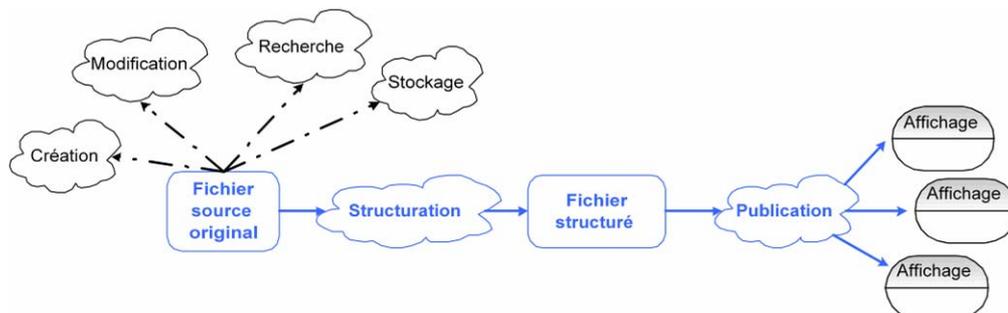


Figure 4.38 – Représentation des fonctions du traitement du document et plus particulièrement la structuration du document et de sa publication.

⁵⁹ <http://www.sgmlsource.com/> (dernier accès le 10 juin 2006).

Nous nous attachons plus spécifiquement aux techniques du traitement du document qui concernent principalement la structuration et l’affichage du document (Figure 4.38) dans le but qui est le nôtre de structurer le texte autour d’éléments spécifiques comme les recommandations.

La structuration du document repose sur un processus de balisage, comme le décrit Goldfarb (Goldfarb, 1981) qui permet d’ajouter des informations dans le texte, dénommées « balises » qui sont utilisées pour séparer les éléments logiques du document, et pour spécifier les fonctions de traitement à exécuter sur ces éléments. On peut distinguer deux types de balisage : (i) le marquage sémantique des balises consiste à ajouter une information sémantique aux balises ; (ii) le marquage sémantique du contenu consiste à introduire des balises en fonction du contenu sémantique du texte balisé.

4.2.1 Le marquage sémantique des balises

Les premiers travaux portant sur l’ajout d’information sémantique dans les documents ont été réalisés autour de GML (*Generalized Markup Language*) et SGML (*Standard Generalized Markup Language*)⁶⁰. SGML est un métalangage structuré normalisé par l’ISO en 1986⁶¹ qui permet de décrire la structure logique d’un document à l’aide d’une DTD (*Document Type Definition*) dans laquelle sont spécifiés la description des éléments, leurs contenus et les relations entre les éléments. Chaque élément d’un document structuré peut alors être stocké, recherché, réutilisé, extrait pour créer un autre document ou une base de données.

Récemment, XML (*eXtended Markup Language*)⁶² a émergé en se basant sur le concept de SGML, mais avec la particularité de personnaliser les balises utilisées (Dymetman, Lux & Ranta, 2000). Les fichiers basés sur XML sont constitués de marquages sémantiques, par exemple le marquage `<description> c’est un guide de bonnes pratiques </description>` n’indique ni la sémantique du contenu (de la portion de texte balisée) ni la manière dont les données doivent être représentées. En d’autres termes, la typologie des balises permet d’identifier dans un document tous les segments de texte correspondant à un concept particulier, mais la sémantique de ce balisage est définie de façon arbitraire. Elle peut correspondre à des indications méta-textuelles (`<definition>`), des catégories correspondant à une ontologie du domaine (`<pathologie>`, `<traitement>`) ou des catégorisations du signifiant (`<valeur numérique>`).

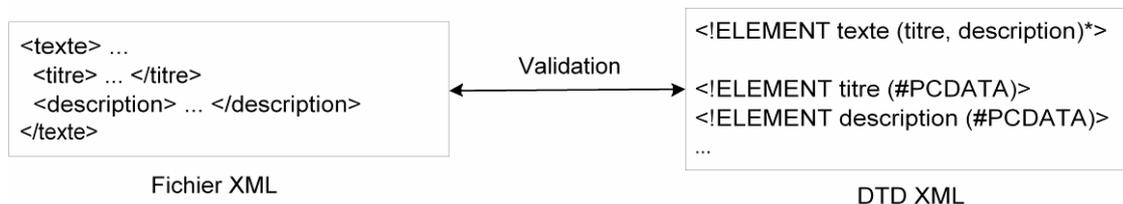


Figure 4.39 – Représentation du contenu d’un fichier XML validé par sa DTD.

⁶⁰ <http://www.sgmlsource.com/> (dernier accès le 10 juin 2006).

⁶¹ <http://www.iso.org/> (dernier accès le 10 juin 2006).

⁶² Recommandation du W3C (*World Wide Web Consortium*) depuis 1998. <http://www.w3.org> (dernier accès le 10 juin 2006).

Les fichiers XML sont définis par une DTD (cf. Figure 4.39), comme cela est également le cas pour SGML. Les données balisées sont indépendantes du média de sortie et ne contiennent aucune indication sur le formatage du document.

Une des fonctionnalités d'un document structuré en XML concerne sa réutilisation et peut donc être interprété pour l'affichage dans plusieurs formats, par exemple HTML, avec les avantages d'affichage, d'accès et de mise en ligne inhérents à ce format. Mais l'avantage principal de la structuration XML dérive de la possibilité d'extraire certains éléments par l'intermédiaire des langages de transformation de documents XML que sont XSLT⁶³ (*eXtensible Stylesheet Language Transformation*) et XQuery⁶⁴ (*an XML Query Language*).

XML permet donc de structurer un document par l'usage de balises et de les personnaliser en fonction de l'application visée. Les analyseurs syntaxiques valident le document XML en vérifiant la conformité des données XML par rapport à la DTD correspondante.

L'interprétation des balises repose quant à elle sur les processeurs XSLT représentant un outil générique pour transformer des documents XML en d'autres documents (par exemple en HTML).

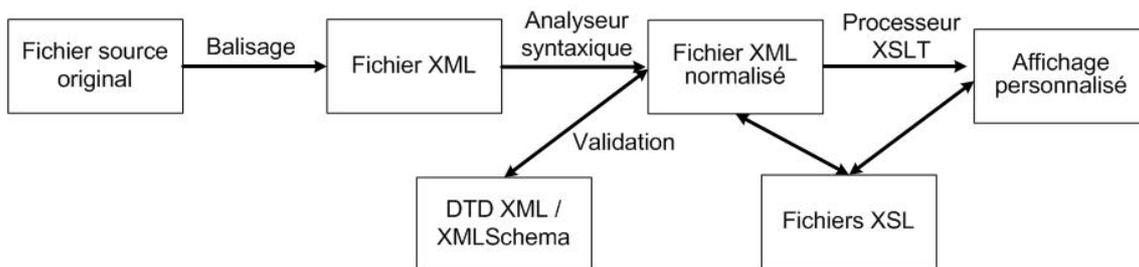


Figure 4.40 – Représentation de la modification d'un fichier en un fichier XML, et du processus d'affichage de ce fichier.

XSLT est basé sur des feuilles de style décrites en XSL⁶⁵ (*eXtensible Stylesheet Language*) et permet l'extraction de certaines données du document XML ainsi que la personnalisation de l'affichage en fonction de ces feuilles de style (Figure 4.40). Ceci offre une flexibilité considérable dans la présentation d'un même document source ; on peut par exemple envisager de définir des niveaux de structuration variables en fonction de catégories d'utilisateurs : il suffit pour cela de définir les feuilles de style XSL correspondantes, qui peuvent aussi évoluer dans le temps, sans pour autant remettre en cause la structuration initiale du document XML. Un autre aspect de ces langages de transformation associés à XML est orienté vers la recherche de sections spécifiques qui sont balisées et qui peuvent être recherchées plus rapidement et de manière plus efficace que dans un document non normalisé (Akhtar, Reilly & Dunnion, 2003).

⁶³ <http://www.w3.org/TR/xslt> (dernier accès le 10 juin 2006).

⁶⁴ <http://www.w3.org/TR/xquery/> (dernier accès le 10 juin 2006).

⁶⁵ <http://www.w3.org/Style/XSL/> (dernier accès le 10 juin 2006).

D'autres langages ont été définis sur la base de XML et sont considérés comme langages de description standard du W3⁶⁶. Par exemple, le format de graphique vectoriel SVG (*Scalable Vector Graphics*), le langage de marquage mathématique (*Mathematical Markup Language* (MathML)), le système RDF (*Resource Description Framework*) pour la description de méta données sur les documents, ou SMIL⁶⁷ (*Synchronized Multimedia Integration Language*) qui permet d'écrire des scénarios pour l'intégration multimédia de texte, image, vidéo et audio. L'exemple ci-dessous illustre l'exécution d'un fichier SMIL qui correspond à l'exécution d'une séquence vidéo présentant SMIL.⁶⁸

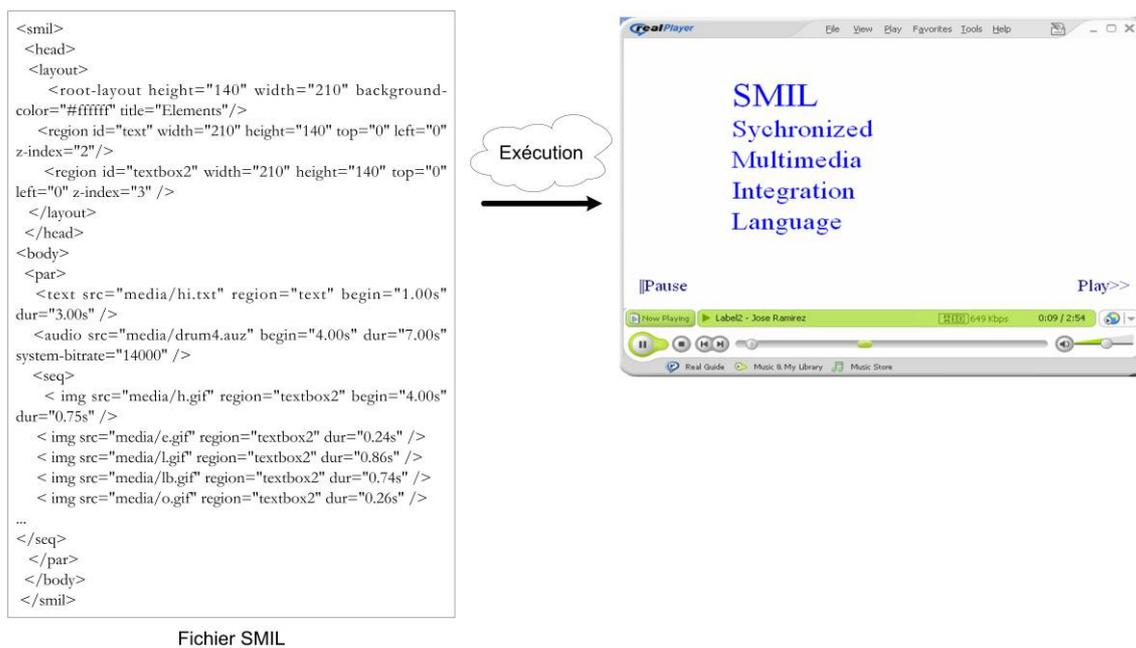


Figure 4.41 – Exécution d'un fichier SMIL l'interprétant comme une séquence vidéo.

Les approches documentaires basées sur XML ont également été introduites pour structurer des documents médicaux et plus particulièrement pour proposer des standards de structuration des Guides de Bonnes Pratiques comme GEM (Shiffman et al., 2000). Ce modèle documentaire très connu dans l'informatisation des Guides de Bonnes Pratiques a été détaillé dans le chapitre 2, et il n'est donc pas nécessaire de détailler sa structure dans ce chapitre. Soulignons simplement, par rapport à la discussion précédente sur le balisage XML, que GEM a su exploiter la puissance, mais dans une certaine mesure aussi l'ambiguïté inhérente à la définition de balises XML correspondant à des éléments du domaine d'application. Une façon de remédier à cette ambiguïté et de garantir à terme une meilleure cohérence du marquage serait d'associer le balisage au véritable contenu sémantique du texte balisé : c'est ce point que nous allons maintenant aborder, d'abord sous l'angle du balisage, puis en reprenant l'histoire récente de l'Extraction d'Informations.

⁶⁶ <http://www.w3.org> (dernier accès le 10 juin 2006).

⁶⁷ <http://www.w3.org/AudioVideo/> (dernier accès le 10 juin 2006).

⁶⁸ <http://www.multimedia4everyone.com/> (dernier accès le 10 juin 2006).

4.2.2 Le marquage sémantique du contenu

Le marquage sémantique du contenu a été initié par la TEI (*Text Encoding Initiative*)⁶⁹ pour représenter les textes littéraires et linguistiques pour la recherche, l'enseignement, les bibliothèques, et l'édition. L'objectif consiste à définir un format d'échange, de création, et de stockage de textes annotés, impliquant une normalisation des balises et une formalisation de leur définition utilisant la syntaxe des DTD. La TEI reprend la distinction entre fichier (unité physique) et document (unité logique) et fournit des versions computationnelles des textes. Dans les *Recommandations*⁷⁰, l'encodage n'est pas une *interprétation* mais une *représentation* du texte. Le but est d'obtenir à partir d'un fichier source original un document encodé XML selon les conventions TEI (Figure 4.42). À partir du corpus XML normalisé TEI, des *parseurs* (ou analyseurs syntaxiques) permettent d'extraire des sous corpus et également de convertir le corpus dans plusieurs formats comme représenté par le schéma ci-dessous.



Figure 4.42 – Description de la démarche TEI.

Les *Recommandations* fournissent le moyen de rendre explicites certaines caractéristiques d'un texte, de façon à faciliter le traitement de ce texte. La TEI dans son ensemble fournit une analyse détaillée des composants de noms, des méta-informations détaillées fournissant des données quant aux origines du texte et aux sujets qu'il aborde, des informations sur l'historique des impressions, ou des variantes manuscrites.

Un texte conforme à la TEI comporte un en-tête TEI (balisé comme un élément `<teiHeader>`) et la transcription du texte lui-même (balisé comme un élément `<text>`). Un document TEI simple comporte, au niveau textuel, les éléments suivants :

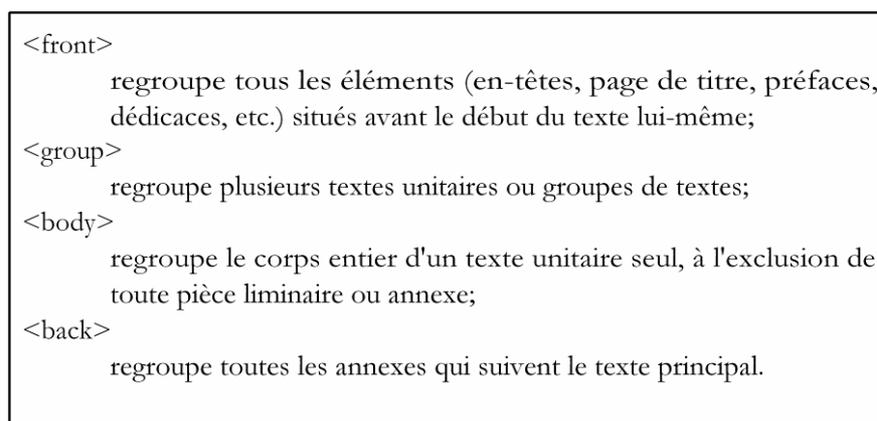


Figure 4.43 – Exemples d'éléments composant un document TEI.

⁶⁹ <http://www.tei-c.org/> (dernier accès le 10 juin 2006).

⁷⁰ <http://www.tei-c.org/P4X/> (dernier accès le 10 juin 2006).

L'en-tête TEI contient une description bibliographique du texte électronique, une description de la manière dont il a été codé, une description non bibliographique du texte (le « profil » du texte) et un historique de révision.

Les fonctionnalités de la TEI sont, par exemple, la distinction des paragraphes (par <p>), le traitement de la ponctuation, par exemple *comma.dec* pour définir la virgule d'un nombre décimal, la distinction des mots ou expressions étrangères représentées par la figure ci-dessous.

<foreign> identifies a word or phrase as belonging to some language other than that of the surrounding text.

<emph> marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

<hi> marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.

<distinct> identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectal, non-preferred, etc., or as forming part of a sublanguage.

Figure 4.44 – Exemples de balises utilisées pour décrire les mots étrangers.

L'exemple ci-dessous représente la reconnaissance et le balisage du mot étranger (à l'anglais) *croissant*.

John eats a <foreign lang="fr">croissant</foreign> every morning.

Figure 4.45 – Extrait d'un fichier normalisé TEI pour lequel la balise *foreign* indique un mot étranger.

D'autres balises TEI sont définies, par exemple, pour décrire les noms, les nombres, les dates, les abréviations, et les adresses. La balise <name> contient un nom propre ou un nom d'une phrase. De la même manière, des balises sont définies pour représenter la date, le temps, les graphes, les réseaux, les arbres, les tableaux, les formules, les graphiques, ou le genre d'un texte par exemple la prose, les vers, les récits dramatiques.

La TEI intègre également la notion de surlignage par l'utilisation de caractéristiques typographiques (la police, la taille, la couleur, etc.) pour distinguer certains passages du texte en contexte. Le but du surlignage permet généralement d'attirer l'attention du lecteur sur certains passages du texte⁷¹.

La TEI permet de baliser des textes de huit genres tels que la prose, la poésie (versifiée), le théâtre, les transcriptions de l'oral, le dictionnaire, la terminologie ; à quoi s'ajoutent deux genres mixtes. D'après Rastier (Rastier, 2000), « Outre qu'elle [la TEI] marque une reconnaissance du problème de la textualité par toutes les communautés (littéraires, linguistiques, informatiques) des traitements automatiques du langage, elle requiert une

⁷¹ Le surlignage et son réel impact sur le lecteur relève des travaux de sciences cognitives qui seraient l'objet d'une thèse.

réflexion nouvelle sur le problème des critères typologiques. ». La TEI propose donc des catégories sémantiques arbitraires par exemple sur le style ou les sentiments mais ces catégories, même arbitraires, ont introduit l'idée d'un balisage se référant à une typologie sémantique du contenu du segment de texte balisé.

Bien que les balises utilisées par la TEI puissent parfois faire référence au contenu du texte balisé, leur association à ces contenus demeure entièrement manuelle. Les noms propres peuvent être signalés par des balises spécifiques mais le balisage ne procède pas à la reconnaissance automatique des noms propres. La reconnaissance automatique du texte correspondant au contenu des balises relève des tâches de l'Extraction d'Informations, et ce sont des systèmes de ce type qui ont les premiers proposés le balisage de sections du texte à partir de la reconnaissance de leur signification (à travers en particulier la reconnaissance de descriptions ou *Named Entity Recognition*). Ces techniques sont exposées dans la section suivante.

4.3 Les techniques de l'Extraction d'Informations

Le balisage traditionnel ajoute une information au texte, alors que le balisage sémantique du contenu s'intéresse à la sémantique des segments textuels balisés ; dans ce dernier les balises délimitent un segment de texte à partir de son contenu, ce qui ouvre la voie à un balisage automatique à partir de l'identification du sens de certaines descriptions caractéristiques.

On désigne sous le terme d'Extraction d'Informations (*Information Extraction*) une approche du traitement automatique des textes qui est basée sur l'identification de contenus sémantiques spécifiques dans un domaine précis (des acteurs, des types d'événements, des lieux...), à partir de l'analyse linguistique du texte, dans le but d'instancier à partir des documents des représentations typiques de certains événements (comme des attentats ou des fusions-acquisitions d'entreprises). L'Extraction d'Informations est basée sur des techniques de traitement automatique du langage naturel souvent empiriques, et fait une part importante à la reconnaissance du contenu de certains segments de phrase, comme dans la reconnaissance de descriptions ou *Named Entity Recognition* (NER).

L'Extraction d'Informations a connu un essor particulier grâce à des évaluations compétitives de la performance des systèmes d'Extraction d'Informations sur les mêmes jeux de test. Ces compétitions sont connues sous l'acronyme de MUC (*Message Understanding Conference*) (DARPA Software and Intelligent Systems Technology Office, 1992), et ont été initiés par la DARPA (*Defense Advanced Research Projects Agency*), une des principales agences de financement des recherches dans ce domaine aux USA. Un exemple de balisage est représenté par la Figure 4.46.

Les systèmes d'Extraction d'Informations utilisent différentes combinaisons de techniques de traitement automatique du langage naturel (TALN), par exemple les automates à états finis, les modèles statistiques. Pour faciliter le traitement de phrases

parfois longues, ces systèmes ont développé des stratégies empiriques de segmentation des phrases, ainsi que de traitements à plusieurs niveaux (mots, segments de phrases, phrases entières).

GARCIA ALVARADO, 56, WAS KILLED WHEN A BOMB PLACED BY URBAN GUERILLAS
ON HIS VEHICLE EXPLODED AS IT CAME TO A HALT AT AN INTERSECTION IN
DOWNTOWN SAN SALVADOR.



```
<\name> GARCIA ALVARADO <\endname>, <\num> 56 {num(56)} <\endnum>,
<cs> WAS {closed(was,[pastv])} <\endcs> <\gls> KILLED
{action(killed,'ATTACK')} <\endgls> <\cs> WHEN
{closed(when,[conj,pron])} <\endcs> <\cs> A {closed(a,[determiner])}
<\endcs> <\weapon> BOMB {type(['BOMB'])} <\endweapon> <\res> PLACED
{atom(placed)} <\endres>
<\cs> BY {closed(by,[prep])} <\endcs> <\res> URBAN {atom(urban)} <\endres>
<\organ> GUERILLAS {type(['TERRORIST','NOUN'])} <\endorgan> <\cs>
ON {closed(on,[prep])} <\endcs> <\cs> HIS
{closed(his,[determiner,pron])} <\endcs> <\target> VEHICLE
{type(['TRANSPORT VEHICLE'])} <\endtarget> <\gls> EXPLODED
{action(exploded,'BOMBING')} <\endgls> <\cs> AS
{closed(as,[conj,pron,prep])} <\endcs> <\cs> IT {closed(it,[pron])}
<\endcs> <\res> CAME {atom(came)} <\endres>
<\cs> TO {closed(to,[prep])} <\endcs> <\cs> A {closed(a,[determiner])}
<\endcs> <\res> HALT {atom(halt)} <\endres>
<\cs> AT {closed(at,[prep])} <\endcs> <\cs> AN {closed(an,[determiner])} <\endcs>
<\res> INTERSECTION {atom(intersection)} <\endres> <\cs> IN {closed(in,[prep])}
<\endcs> <\res> DOWNTOWN {atom(downtown)} <\endres> <\place> SAN SALVADOR
{type(['CITY','EL SALVADOR'],['DEPARTMENT','EL SALVADOR'])} <\endplace>.
```

Figure 4.46 – Exemple de balisage dans le système *MucBruce* (Cowie, Guthrie & Wilks, 1992).

Si les systèmes d'Extraction d'Informations adoptent des approches très diverses quant aux techniques qu'ils assemblent, ils partagent en revanche certaines stratégies qui ont donné lieu à des tâches spécifiques comme la reconnaissance de descriptions (*Named Entity Recognition*). La Figure 4.47 représente l'architecture d'un système d'Extraction d'Informations typique, le système Proteus/PET (Yangarber & Grishman, 1998).

En général, pour chaque étape d'un système d'Extraction d'Informations, une technique de TALN est invoquée, produisant un assemblage parfois complexe de techniques hétérogènes. Néanmoins, dans l'histoire des systèmes d'Extraction d'Informations, le système FASTUS s'est distingué par sa démarche empirique entièrement basée sur une cascade d'automates à états finis (Hobbs et al., 1996). Un niveau pouvait être consacré à l'identification de descriptions spécifiques, par exemple les acteurs du récit, puis à d'autres comme à l'attribution de rôles à ces différents acteurs dans l'action reconnue à partir du texte. FASTUS a en particulier prouvé son efficacité pour la reconnaissance d'acteurs à partir de descriptions en langage naturel.

Mais l'Extraction d'Informations a aussi innové dans l'utilisation du balisage au cours du traitement automatique : dès 1992, c'est-à-dire bien avant l'essor des techniques documentaires autour de XML, les systèmes de ce type ont utilisé le balisage en cours

d'analyse pour signaler l'identification d'une structure linguistique particulière, comme un nom propre, un lieu, etc.

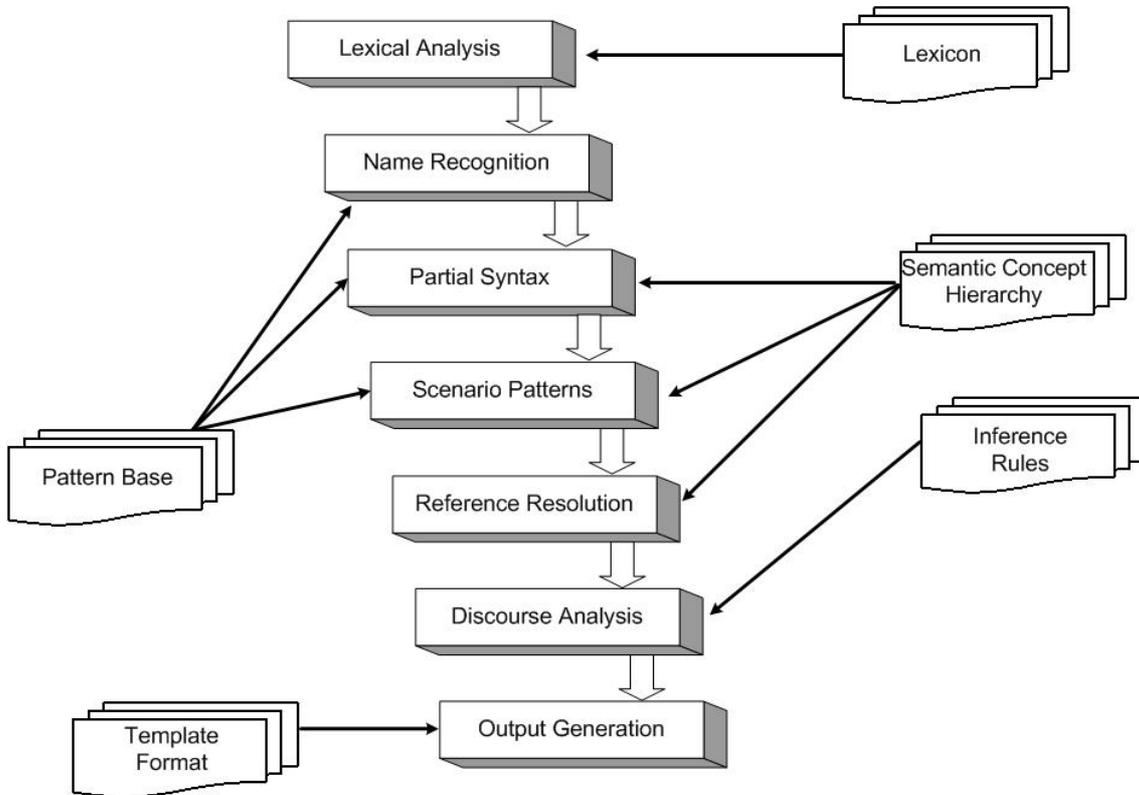


Figure 4.47 – Architecture d'un système typique d'Extraction d'Informations (Proteus/PET) d'après Yangerber et al. (Yangerber & Grishman, 1998).

Cette approche de la reconnaissance de certains motifs a donné naissance à la tâche dite NER (Hirschman & Chinchor, 1998), en tant que sous-tâche de l'Extraction d'Informations. Les tâches de NER consistent en trois sous-tâches : les entités nommées (organisations, personnes, lieux), les expressions temporelles (dates, temps), et les expressions numériques (valeurs monétaires, pourcentages). Par exemple, un système NER produit dans le style MUC la sortie suivante :

```

Jim bought 300 shares of Acme Corp. in 2006.

<ENAMEX TYPE="PERSON">Jim </ENAMEX> bought<NUMEX
TYPE="QUANTITY">300</NUMEX> shares of<ENAMEX
TYPE="ORGANIZATION">Acme Corp.</ENAMEX> in<TIMEX
TYPE="DATE">2006</TIMEX>.
    
```

Figure 4.48 – Balisage d'une phrase par un système NER selon la représentation MUC.

Les systèmes NER permettent la reconnaissance de structures nominales ; nous avons choisi de nous inspirer de cette approche pour la reconnaissance automatique des opérateurs déontiques, qui sont des structures verbales. Dans les deux cas, il s'agit de

reconnaître des motifs à partir de multiples présentations lexicales et syntaxiques. Les structures syntaxiques des opérateurs déontiques tendent à être plus complexes en tant que structures verbales (insertion d'adverbes, constructions actives et passives, impersonnelles ...), ce qui justifie bien le recours à des techniques d'analyse.

Nous proposons une approche consistant à fusionner le traitement linguistique et le traitement documentaire. Nous décrivons dans les chapitres suivants la mise en œuvre de notre approche pour reconnaître les opérateurs déontiques. Nous avons choisi d'utiliser les automates à états finis (Roche & Schabes, 1997) comme formalisme de traitement du fait qu'ils fournissent une façon efficace pour définir les structures spécialisées, en incluant leurs variantes morphologiques, et ont une performance prouvée sur certaines tâches d'Extraction d'Informations (Pazienza, 1997) (tel que la reconnaissance de descriptions), qui sont similaires à la reconnaissance de structures déontiques. L'application d'un automate reconnaît une occurrence spécifique d'un opérateur déontique dans le texte et peut délimiter cette occurrence à l'aide de balises qui serviront aussi à la structuration du texte.

Pour développer des automates, de nombreux logiciels existent tels qu'Unitex⁷², ou, plus récemment, des *Recommandations* du W3C ont été proposées pour définir des automates : le *State Chart XML*⁷³ (SCXML) pour *State Machine Notation for Control Abstraction 1.0* (version de Juillet 2005). Cette technique aurait également pu être choisie si le développement de nos automates n'avait pas été antérieur à ces *Recommandations* du W3C.

Néanmoins, nous avons opté pour un développement spécifique qui permet de contrôler l'intégration du système dans un environnement graphique et ne limite pas les utilisations ultérieures du logiciel, y compris son exploitation et sa valorisation.

⁷² <http://www-igm.univ-mlv.fr/~unitex/> (dernier accès le 10 juin 2006).

⁷³ <http://www.w3.org/TR/2005/WD-scxml-20050705/> (dernier accès le 10 juin 2006).

Chapitre 5

L'environnement G-DEE

Ce chapitre décrit l'environnement G-DEE (Guideline Document Engineering Environment) que nous avons développé. Il s'attache à l'interface graphique, l'utilisation et aux fonctionnalités de l'approche présentée dans cette thèse. G-DEE est dédié à la fois à l'étude des Guides de Bonnes Pratiques et à terme à leur analyse automatique, et intègre de nombreuses fonctionnalités que nous décrivons à travers ses modes d'utilisation.

5.1 Description générale de G-DEE

G-DEE pour *Guideline Document Engineering Environment*, est l'environnement graphique que nous avons développé et dédié à l'étude des Guides de Bonnes Pratiques. La philosophie de G-DEE représente une extension des interfaces graphiques actuelles pour l'encodage des Guides de Bonnes Pratiques, comme GEM-Cutter⁷⁴. L'innovation principale de cet environnement consiste à intégrer des fonctions de traitement automatique des textes qui permettent l'identification automatique d'éléments structurants du texte, comme les recommandations élémentaires. L'intégration de ces fonctions intelligentes dans une interface graphique augmente considérablement la capacité de l'utilisateur à explorer et manipuler les documents.

Un analyseur basé sur des automates à états finis repère les expressions des opérateurs déontiques à l'intérieur des Guides de Bonnes Pratiques, produisant ainsi une première version structurée du Guide de Bonnes Pratiques (Georg & Jalent, 2005). À partir de cette première représentation, les recommandations des Guides de Bonnes Pratiques sont automatiquement balisées (cf. Figure 5.49 partie B).

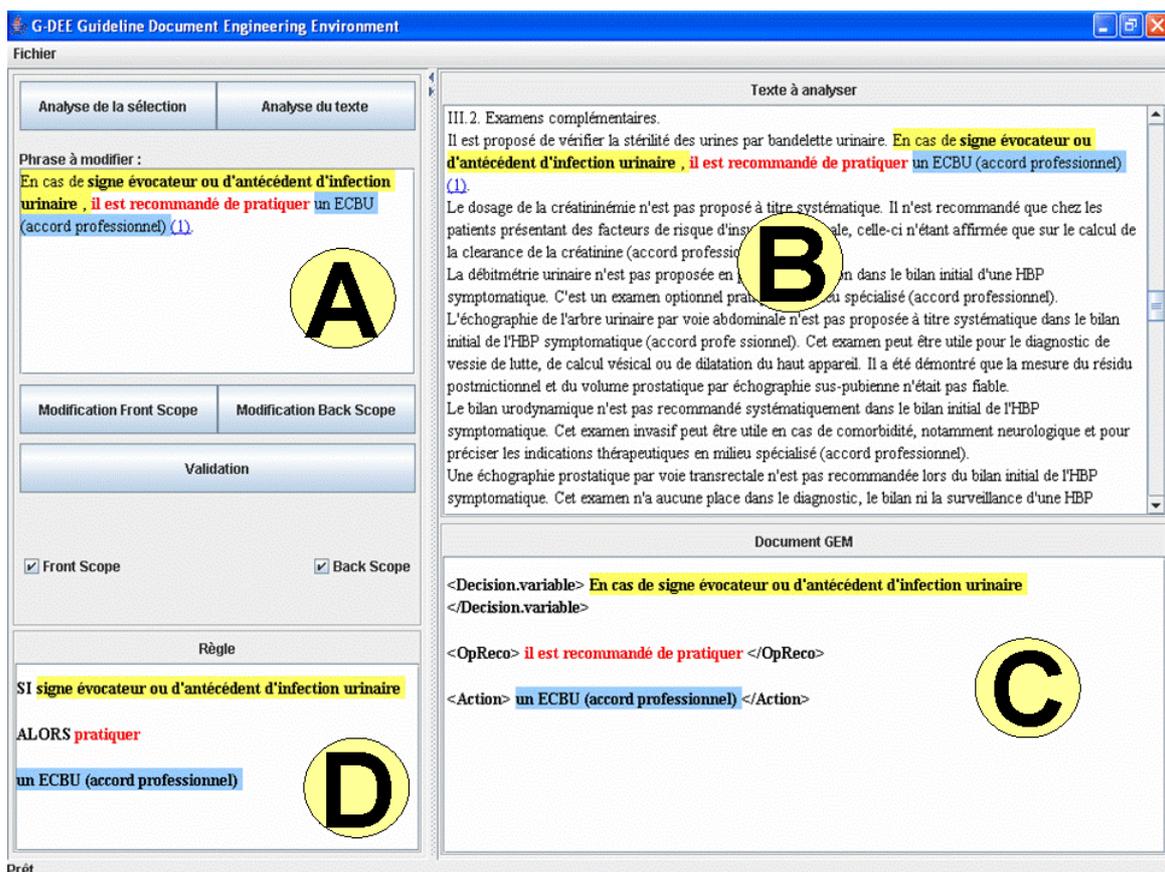


Figure 5.49 – Vue d'ensemble de l'environnement G-DEE.

⁷⁴ <http://gem.med.yale.edu/> (dernier accès le 10 juin 2006)

À partir de ce balisage, il est possible d'identifier les éléments tels que les variables de décision, de produire des éléments encodés en GEM (cf. Figure 5.49 partie C), et de produire des règles de décision (cf. Figure 5.49 partie D) de façon automatique en utilisant des procédures d'extraction et de transformation du contenu balisé.

5.2 Les zones d'affichage de G-DEE

5.2.1 Analyse et structuration du texte

La structuration des Guides de Bonnes Pratiques est basée sur la reconnaissance automatique des opérateurs déontiques appartenant aux recommandations. Le texte est analysé puis structuré autour des occurrences d'opérateurs déontiques. Les structures ainsi obtenues sont présentées visuellement en utilisant un code de couleur. Les phrases contenant les opérateurs déontiques sont surlignées : en rouge pour l'opérateur déontique, en jaune pour le *front-scope* et en bleu pour le *back-scope* (Figure 5.50)⁷⁵. Cette représentation est destinée à être plus lisible qu'un fichier XML, en particulier, dans le futur, pour des utilisateurs non-informaticiens.

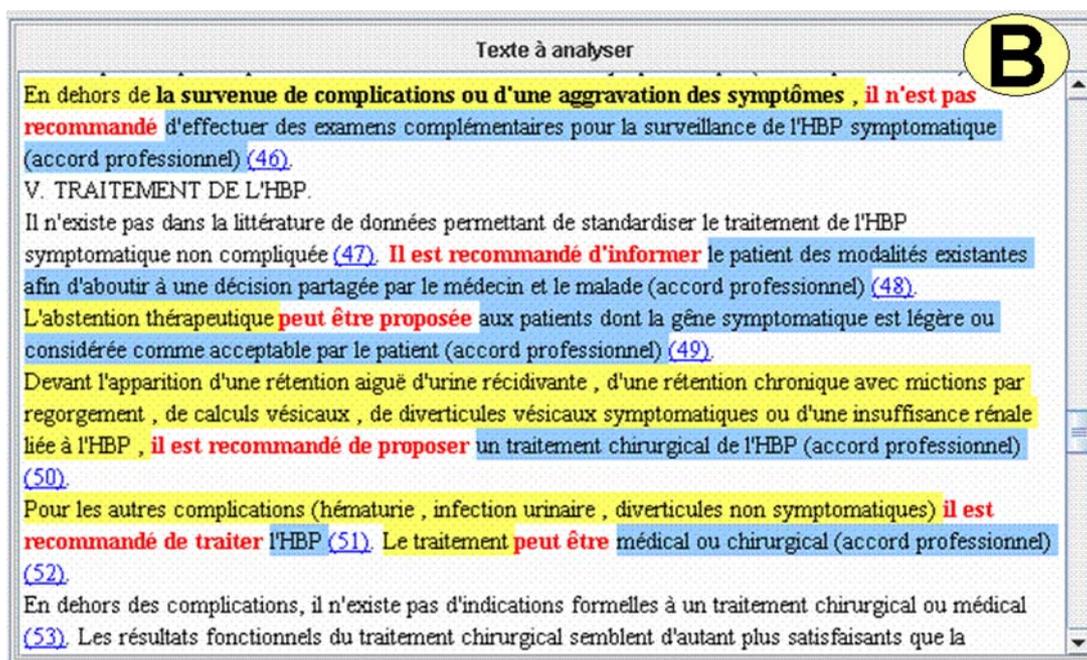


Figure 5.50 – Présentation du balisage des opérateurs déontiques et de leurs *scopes* associés dans le texte.

⁷⁵ Le *front-scope* est la partie textuelle qui précède l'opérateur déontique, et le *back-scope* est la partie textuelle qui suit l'opérateur déontique. Nous reviendrons sur ces notions dans le chapitre 6.

5.2.2 Structuration selon le modèle GEM

La reconnaissance d'opérateurs déontiques et donc de recommandations élémentaires peut servir de point de départ à une structuration plus fine, en particulier de type GEM (qui a été un des formalismes privilégiés de ces travaux). Une phrase correspondant à une recommandation peut ainsi être analysée pour identifier les éléments correspondant aux variables de décision et aux actions du modèle GEM (Georg, Séroussi & Bouaud, 2005; Shiffman et al., 2000). La phrase est ensuite affichée (Figure 5.51) selon un ensemble de règles définies dans le chapitre 7. G-DEE peut ainsi servir d'environnement semi-automatique pour l'encodage au standard GEM, offrant des fonctionnalités plus avancées que celles de GEM-Cutter.

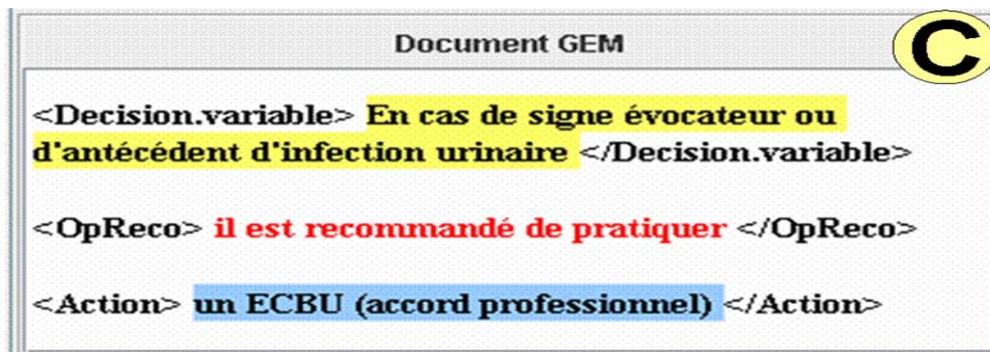


Figure 5.51 – Dérivation automatique des variables de décision et d'action dans le modèle GEM à partir du balisage des opérateurs déontiques et de leurs *scopes* associés.

5.2.3 Génération de règles de décision

Le processus pour identifier les parties correspondantes à la prémisse et à l'action d'une règle de décision comme représenté par la figure ci-dessous repose sur le même processus que celui établi pour dériver de manière automatique les éléments GEM à partir du texte. G-DEE permet alors de visualiser des règles de décision correspondant aux segments de texte sélectionnés, ce qui permet une exploration et une analyse du texte de recommandations directement sous forme de règles SI-ALORS. En dehors des aspects ergonomiques d'assistance à la lecture, cette fonctionnalité peut aider à l'acquisition de connaissances pour des systèmes d'aide à la décision à base de règles.

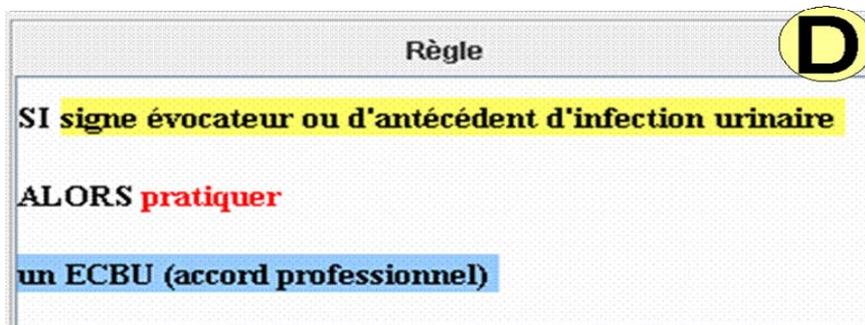


Figure 5.52 – Dérivation automatique des prémisses et des actions d'une règle de décision à partir du balisage des opérateurs déontiques et de leurs *scopes* associés.

5.3 Aperçu des fonctionnalités

L'utilisation de l'interface repose sur une série de fonctions interactives de manipulation du texte. Pour apparaître naturelles, elles suivent les conventions les plus connues et reposent sur des sélections de portions de texte (à l'aide de la souris et du curseur) et l'activation des boutons de l'interface sur les segments sélectionnés. La valeur ajoutée de l'interface découle naturellement des fonctions intelligentes activables de façon interactive.

5.3.1 Analyse d'une sélection ou du texte entier

L'analyseur peut être lancé soit sur la sélection d'une phrase (effectuée préalablement) soit sur l'ensemble du texte (Bouton 1 de la Figure 5.53).

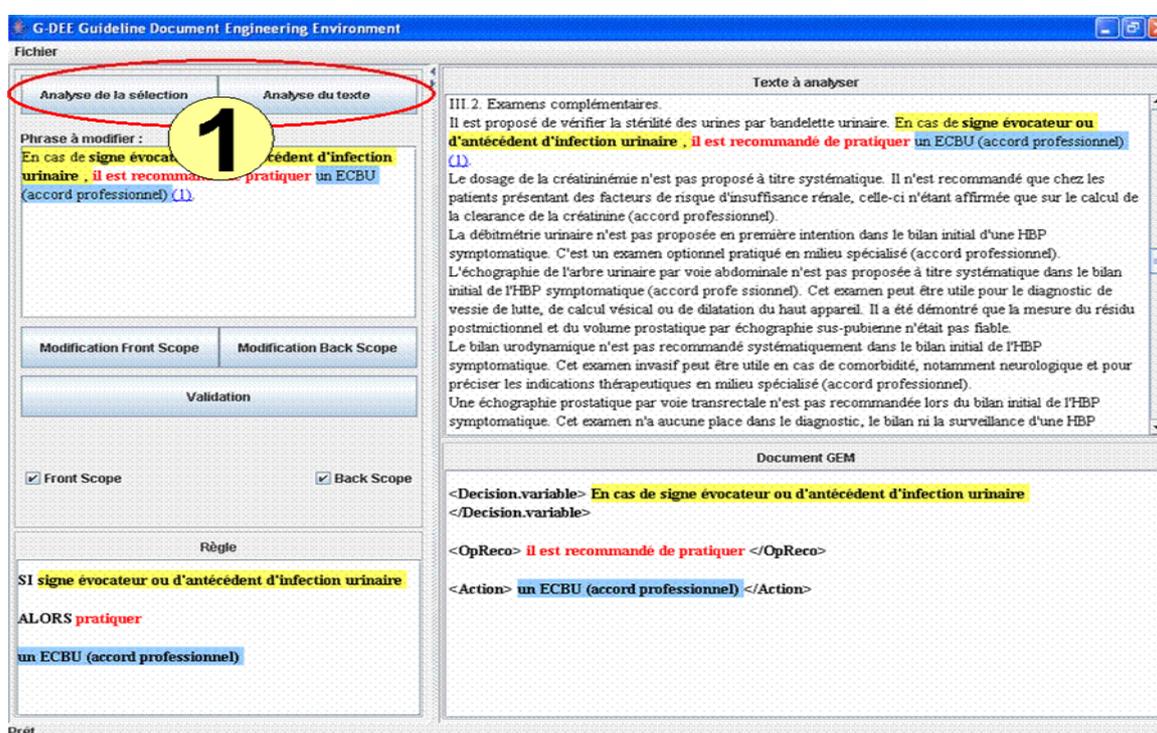


Figure 5.53 – Deux fonctionnalités, analyse d'une sélection ou analyse sur tout le texte.

5.3.2 Affichage *Front-Scope* et/ou *Back-Scope*

Il est possible d'afficher les recommandations structurées à différents niveaux de détail. Les structures complètes correspondent aux opérateurs déontiques reconnus dans le texte sélectionné et analysé, ainsi qu'à leur *front-scope* et *back-scope* (Boutons 3 la Figure 5.54), également surlignés.

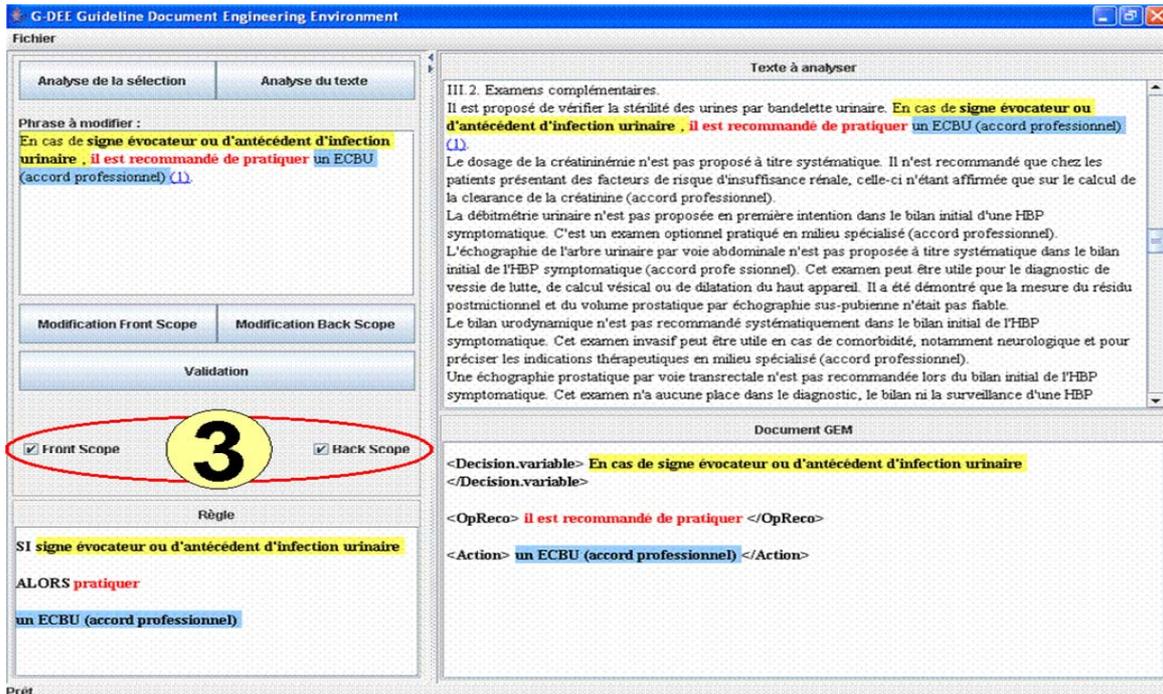


Figure 5.54 – Affichage du *front-scope*, de l'opérateur déontique et du *back-scope*.

Une fois le principe de structuration acquis par l'utilisateur, il peut être utilisé pour une étude plus fine des différentes parties des recommandations. La Figure 5.55 montre l'affichage de l'opérateur et du *front-scope*.

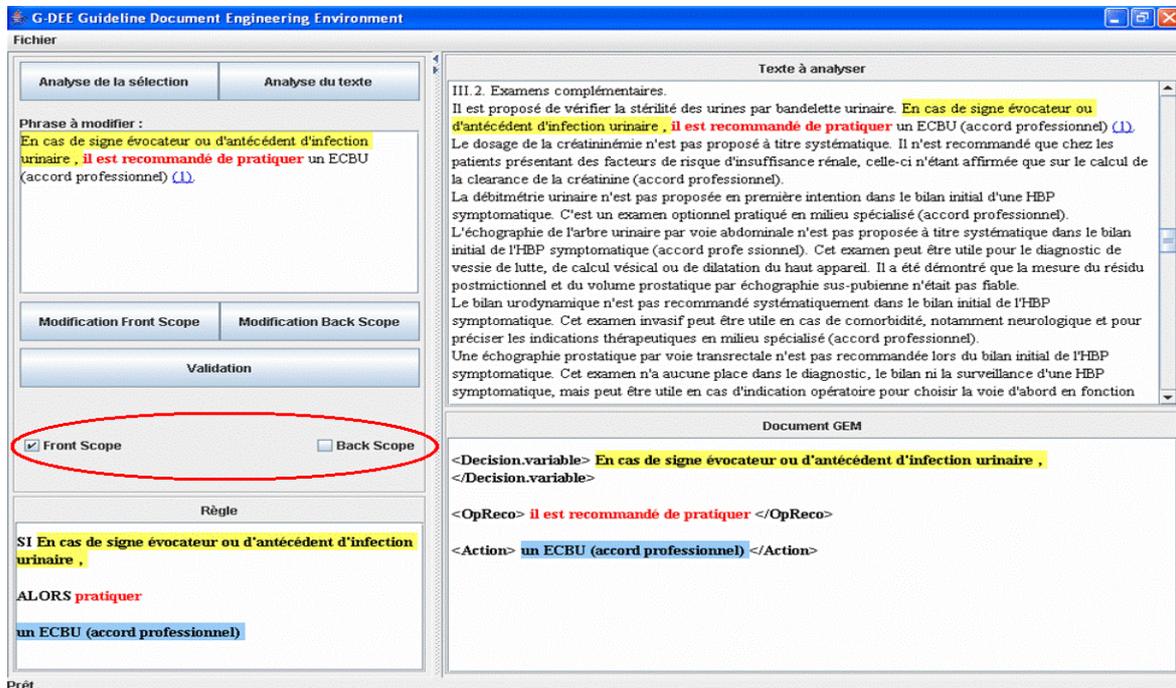


Figure 5.55 – Affichage du *front-scope* et de l'opérateur déontique.

De la même façon, la Figure 8 affiche exclusivement les opérateurs déontiques et leur *back-scope* (Figure 5.56). Ce type d'affichage permet de parcourir les textes à la recherche des

5.3. Aperçu des fonctionnalités

actions diagnostiques ou thérapeutiques faisant partie des recommandations. Le maintien de l'opérateur déontique facilite la lisibilité.

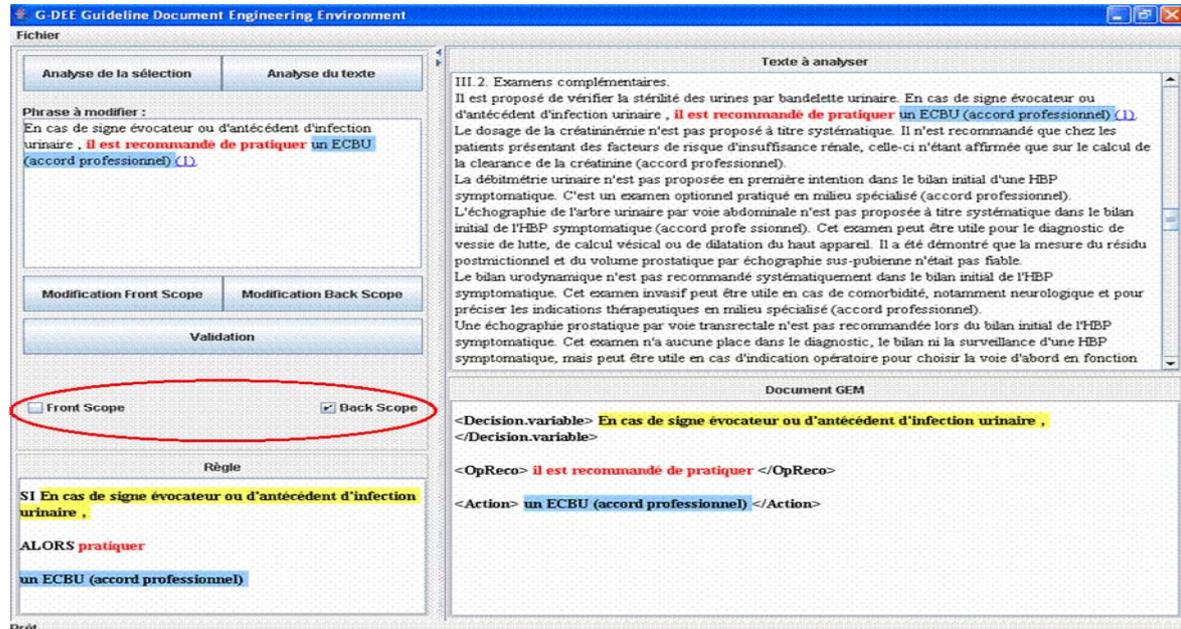


Figure 5.56 – Affichage de la phrase selon deux des trois composants de structuration : l'opérateur déontique et le *back-scope*.

Enfin, il est également possible d'afficher uniquement les opérateurs déontiques dans le texte sélectionné et analysé (Figure 5.57). La Figure 5.58 illustre un texte ainsi balisé, pour lequel les opérateurs déontiques apparaissent en rouge.

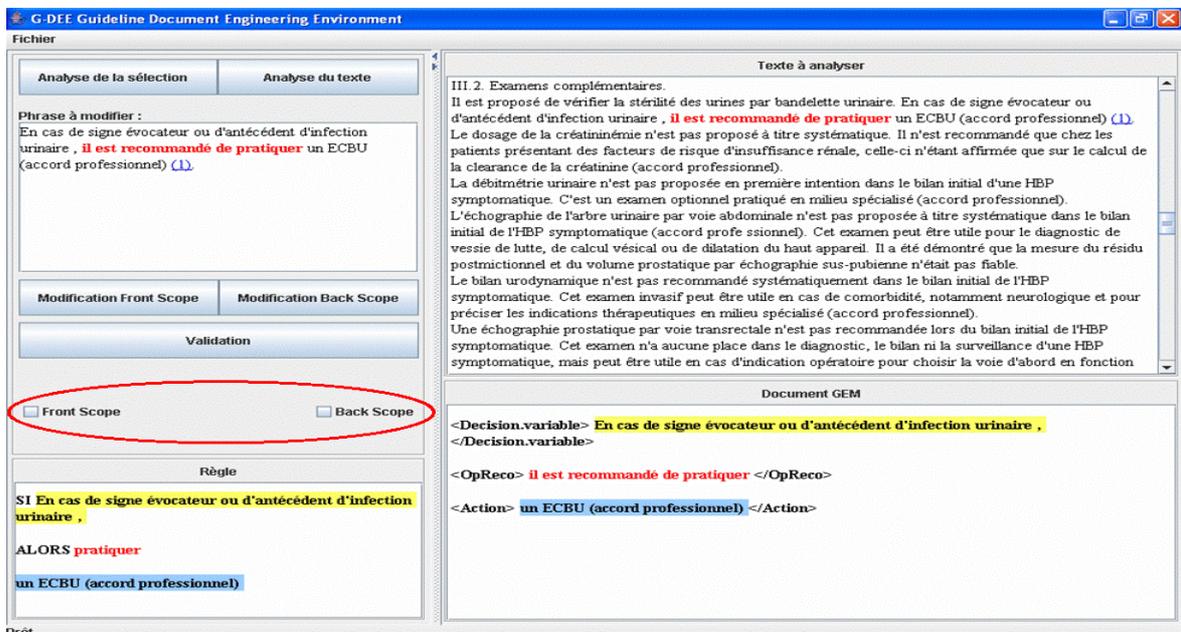


Figure 5.57 – Affichage de l'opérateur déontique seul.

Ce type d'affichage peut servir pour les analyses de distribution, ainsi que (en phase de développement) pour la mise au point et l'évaluation de l'analyseur.

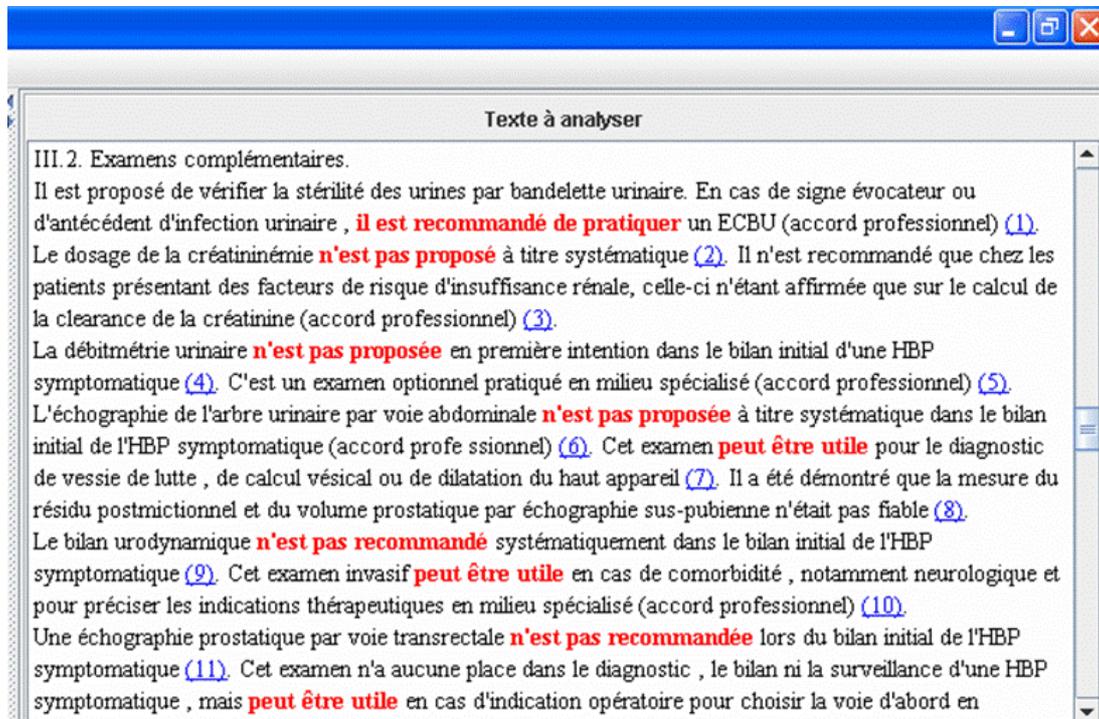


Figure 5.58 – Affichage des opérateurs déontiques seuls du texte.

5.3.3 Modification / Validation d'une phrase

Une partie de l'interface est réservée à la modification et à la validation d'une phrase (Figure 5.59), qu'elle soit reconnue ou non comme une recommandation (en phase de développement).

Cette modification interactive permet d'homogénéiser le style rédactionnel et de lutter contre les ambiguïtés. Par exemple, certaines assertions reconnues comme des recommandations par des experts mais n'incorporant pas d'opérateurs déontiques lors de leur formulation initiale peuvent être modifiées pour faire apparaître un opérateur déontique (en vérifiant interactivement sa reconnaissance par le système).

5.3. Aperçu des fonctionnalités

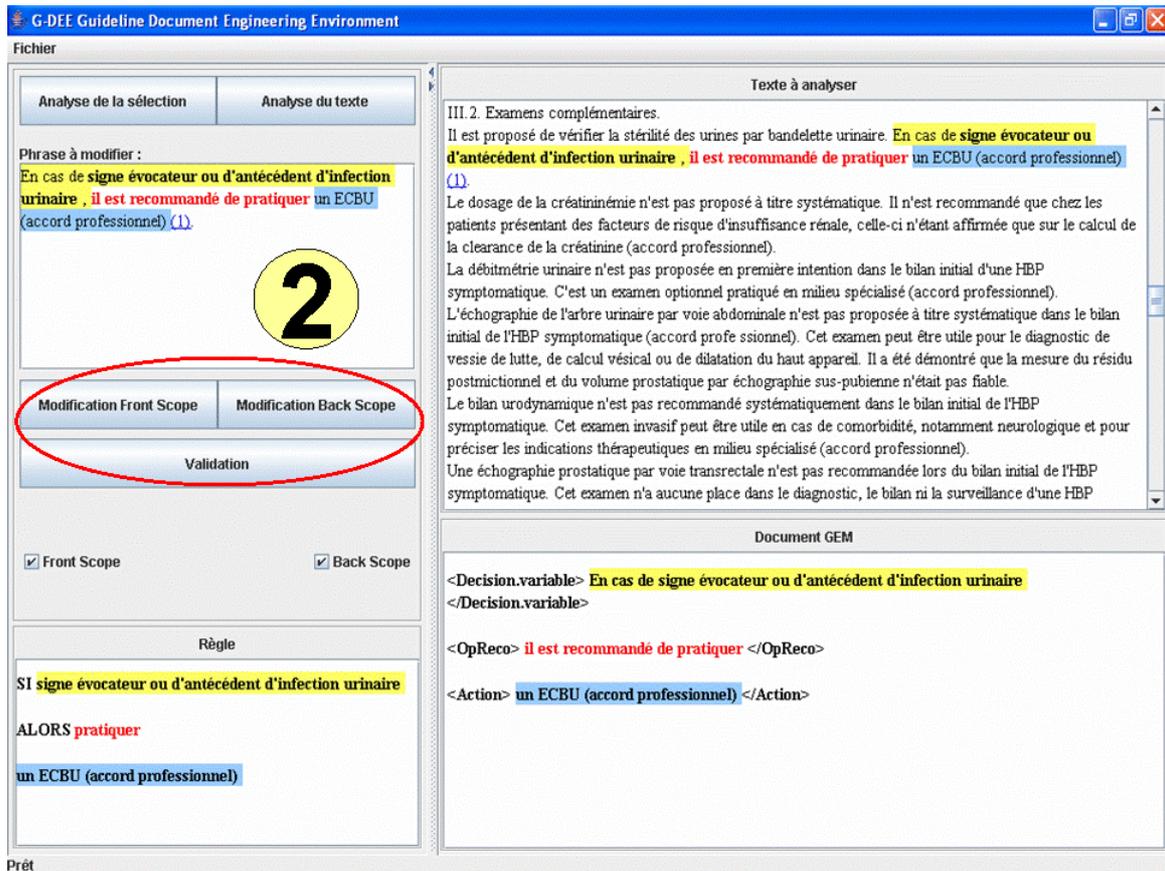


Figure 5.59 – Module de modification et de validation d'une phrase.

Chapitre 6

Techniques d'analyse linguistique

Ce chapitre est dédié aux techniques de traitement linguistique, plus spécifiquement à la description des automates à états finis qui permettent de reconnaître de manière automatique les marqueurs textuels décrits dans le chapitre 3. Ce chapitre intéressera les lecteurs attachés aux aspects informatiques les plus techniques de G-DEE (c'est-à-dire les formalismes d'analyse). Dans une première étape, nous décrivons les pré-traitements nécessaires à l'analyseur, puis nous décrivons les automates à états finis permettant la reconnaissance des marqueurs textuels.

6.1 Stratégie de reconnaissance de motifs

6.1.1 Reconnaissance des opérateurs déontiques

Moulin et Rousseau (Moulin & Rousseau, 1990) ont les premiers montré que l'identification automatique des opérateurs déontiques dans les textes légaux pouvait servir à l'acquisition des connaissances. Nous avons proposé de transposer et d'étendre cette approche au cas des Guides de Bonnes Pratiques cliniques en utilisant la reconnaissance d'opérateurs déontiques ainsi que leur portée pour identifier les recommandations et structurer le document autour de celles-ci.

Dans cette section, nous décrivons les différentes étapes de la reconnaissance automatique des opérateurs déontiques dans les Guides de Bonnes Pratiques. La première étape consiste en un pré-traitement des phrases à analyser pour faciliter l'application de techniques de TAL identifiant les opérateurs déontiques. Ce pré-traitement permet la segmentation de phrases longues ou comprenant plusieurs opérateurs déontiques. La deuxième étape consiste à rechercher le motif de l'opérateur déontique en utilisant une base d'automates à états finis. Cette étape se décompose en cinq modules : (i) sélection d'un motif ; (ii) instanciation du motif sélectionné ; (iii) concordance du motif avec la phrase analysée ; (iv) sélection du motif le plus concordant, i.e. la taille du motif et le nombre d'éléments concordant entre le motif et la phrase analysée⁷⁶ ; (v) identification de la voix active ou passive de l'opérateur déontique. La troisième étape consiste à baliser l'opérateur déontique en fonction du motif concordant.

L'exemple ci-dessous permet d'illustrer ces étapes de reconnaissance et de balisage de l'opérateur déontique.

*« Si le risque est faible ou moyen, **il est recommandé de débiter** un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque. »*

La première étape correspond au pré-traitement segmentant la phrase en deux propositions du fait que le nombre d'occurrences de l'auxiliaire « être » est supérieur au nombre d'occurrences de verbes au participe passé. La première proposition correspond donc à :

« Si le risque est faible ou moyen, »

et la seconde à :

⁷⁶ Cette étape est rendue nécessaire par l'existence de structures partagées entre automates à états finis : lorsque plusieurs automates peuvent s'instancier sur un même segment de phrase il importe de sélectionner le motif donnant le recouvrement maximal (le plus spécifique).

6.2. Règles de description syntaxique

« **il est recommandé de débiter** un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque. »

La deuxième étape consiste en l'identification des motifs concordants. Pour la première proposition, aucun motif n'est reconnu. Cette proposition ne contient pas d'opérateur déontique. Pour la seconde proposition, deux motifs peuvent être concordants :

- (i) l'un correspondant à la reconnaissance de la voix active impersonnelle de *recommander* (3 éléments du motif sur 3 sont concordants, i.e. pronom, auxiliaire *être* suivi du participe passé, avec une taille de motif égale à 3) représenté en rouge dans l'extrait ci-dessous :

« **il est recommandé** ... »

- (ii) l'autre correspondant à la reconnaissance de la voix active impersonnelle (*recommander*) et du verbe (*débiter*) à l'infinitif (5 éléments du motif sur 5 sont concordants, i.e. pronom, auxiliaire être suivi du participe passé, d'une préposition et d'un verbe à l'infinitif, avec une taille de motif égale à 5) représenté en rouge ci-dessous :

« **il est recommandé de débiter** ... »

Le second motif est alors retenu pour le processus de balisage de l'opérateur déontique du fait qu'il est plus spécifique que le premier.

La troisième étape représentée par la figure ci-dessous consiste à baliser l'opérateur déontique en fonction du motif retenu.

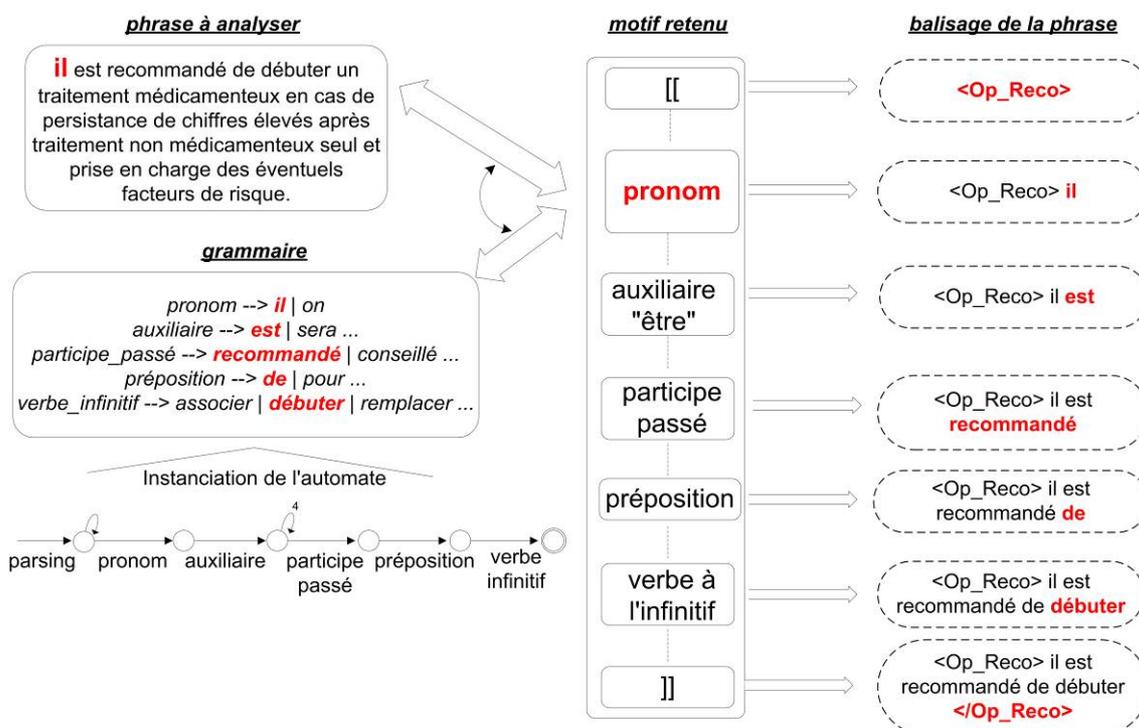


Figure 6.60 – Processus de balisage de l'opérateur déontique en fonction du motif.

Le formalisme⁷⁷ utilisé pour décrire les motifs permet de définir un marqueur de début de balisage de l'opérateur déontique représenté par « [[» dans le motif et correspondant à l'ajout de *<Op_Reco>*⁷⁸ dans la phrase balisée.

Le marqueur de fin de balisage est représenté par «]] » dans le motif et correspond à l'ajout de *</Op_Reco>* dans la phrase balisée. L'élément du motif représentant le pronom est instancié à partir de la grammaire. La concordance entre le pronom instancié et la phrase analysée permet d'instancier la phrase balisée par le pronom. Le processus est répété à l'identique pour les autres éléments du motif.

La sous-chaîne ainsi balisée est concaténée avec le reste de la phrase, et la deuxième proposition balisée correspond donc à :

<Op_Reco> il est recommandé de débiter </Op_Reco> un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque.

Cette proposition est alors concaténée à la première proposition ne contenant pas d'opérateur déontique. La phrase correctement balisée correspond à :

Si le risque est faible ou moyen, <Op_Reco> il est recommandé de débiter </Op_Reco> un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque.

6.1.2 Reconnaissance des *scopes*

Les *scopes* correspondent aux parties textuelles entourant l'opérateur déontique (Moulin & Rousseau, 1990). Plus précisément, le *front-scope* est la partie textuelle qui précède l'opérateur déontique, et le *back-scope* est la partie textuelle qui suit l'opérateur déontique. D'un point de vue fonctionnel, ils constituent les opérands de l'opérateur déontique.

Dans le but de reconnaître et de baliser les *scopes* de l'opérateur déontique, il est auparavant nécessaire d'effectuer le balisage de l'opérateur déontique. En effet, les marqueurs pour reconnaître les *scopes* sont les balises de début et de fin de l'opérateur déontique, comme représenté par la Figure 6.61.

Le marqueur « [[» correspond au premier élément du motif de balisage des *scopes*. Ce marqueur permet d'indiquer le début du *front-scope*. C'est le premier élément de la phrase balisée. Le marqueur de fin de balise du *front-scope*, i.e. *</Front_Scope>*, est le début de balise de l'opérateur déontique *<Op_Reco>* cela permet de concaténer la partie textuelle précédant l'opérateur déontique « *Si le risque est faible ou moyen,* » à la phrase balisée correspondant à « *<Front_Scope>* ». La balise de fin du *front-scope* « *</Front_Scope>* » est concaténée à la phrase balisée pour obtenir la phrase suivante :

⁷⁷ Dans cette section, nous nous intéressons uniquement au balisage de l'opérateur déontique. Nous spécifions le formalisme des motifs dans la section décrivant le fonctionnement de l'analyseur.

⁷⁸ La balise *<Op_Reco_P>* correspond à un opérateur déontique à la voix passive et la balise *<Op_Reco>* correspond à un opérateur déontique à la voix active.

6.2. Règles de description syntaxique

« **<Front_Scope>** *Si le risque est faible ou moyen,* **</Front_Scope>**
<Op_Reco_P> »

De la même manière, la fin de balise de l'opérateur déontique, i.e. **</Op_Reco>**, indique le début de balisage du *back-scope*, **<Back_Scope>**.

<Front_Scope> *Si le risque est faible ou moyen,* **</Front_Scope>**
<Op_Reco> *il est recommandé de débiter* **</Op_Reco>**
<Back_Scope> *un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque.*

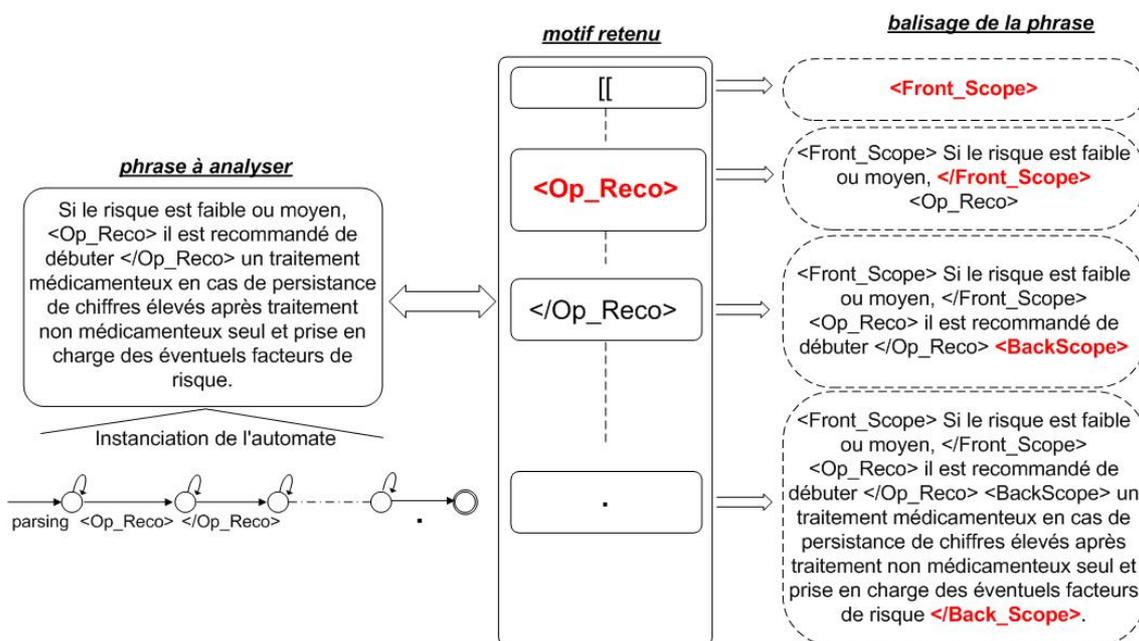


Figure 6.61 – Reconnaissance et balisage des *scopes* de l'opérateur déontique « il est recommandé de débiter ».

La fin de balise du *back-scope* « **</Back_Scope>** » se situe avant le point de ponctuation final comme l'illustre le balisage final de la phrase analysée.

<Front_Scope> *Si le risque est faible ou moyen,* **</Front_Scope>**
<Op_Reco> *il est recommandé de débiter* **</Op_Reco>**
<Back_Scope> *un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque* **</Back_Scope>**.

La reconnaissance et le balisage des *scopes* pour une phrase contenant plusieurs opérateurs déontiques consistent à traiter les deux propositions issues de la segmentation. Par exemple, un pré-traitement est nécessaire pour reconnaître les deux opérateurs déontiques de la phrase ci-dessous.

Une bithérapie orale **est conseillée** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel).

Le pré-traitement segmente la phrase en deux propositions et le module de balisage des opérateurs déontiques est exécuté pour chaque proposition, ce qui permet le balisage de chaque opérateur déontique comme suit :

Proposition 1 : Une bithérapie orale **<Op_Reco_P> est conseillée </Op_Reco_P>** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale,

Proposition 2 : elle **<Op_Reco_P> est recommandée </Op_Reco_P>** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel).

Le module de balisage des *scopes* s'exécute de la même manière que dans l'exemple précédent. En effet, la proposition revient à une phrase contenant un seul opérateur déontique.

Proposition 1 : **<Front_Scope> Une bithérapie orale </Front_Scope> <Op_Reco_P> est conseillée </Op_Reco_P> <Back_Scope>** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, **</Back_Scope>**

Proposition 2 : **<Front_Scope> elle </Front_Scope> <Op_Reco_P> est recommandée </Op_Reco_P> <Back_Scope>** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel) **</Back_Scope>**.

Pour scinder les deux propositions, le module de balisage des *scopes* recherche la proposition contenant le point de ponctuation final. Dans l'exemple présenté ci-dessus, elle correspond à la proposition 2. La proposition 1 est alors concaténée à la proposition 2 pour reconstituer la phrase initiale à analyser. La phrase contenant deux opérateurs déontiques est donc balisée comme suit :

<Front_Scope> Une bithérapie orale </Front_Scope> <Op_Reco_P> est conseillée </Op_Reco_P> <Back_Scope> lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, **</BackScope> <Front_Scope> elle </Front_Scope> <Op_Reco_P> est recommandée </Op_Reco_P>** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel) **</Back_Scope>**.

6.2 Règles de description syntaxique

Comme pour tous les systèmes de TAL, notre analyseur repose sur la définition de ressources linguistiques, sous la forme d'une grammaire décrivant les différents motifs correspondant aux opérateurs déontiques. L'analyseur comprend un algorithme d'analyse et une grammaire, celle-ci étant formalisée en utilisant une série d'automates à états finis. Une des étapes de la construction de l'analyseur consiste donc à définir, à partir d'un corpus de textes représentatifs, la grammaire des opérateurs déontiques. C'est cette méthodologie que nous décrivons à présent.

6.2.1 Composition du corpus

Le corpus que nous avons constitué pour décrire la grammaire des opérateurs déontiques inclus neuf rapports de conférences de consensus (concernant la maladie de Parkinson, la sclérose en plaques, la dépendance aux opiacés, l'hépatite C, les coliques néphrétiques, les épilepsies, et l'urticaire chronique), six cours de médecine (pour les domaines du diabète, de l'hypertension, du risque cardio-vasculaire et de l'insuffisance rénale) et sept Guides de Bonnes Pratiques (concernant la lombalgie chronique, la migraine, la pneumonie, le diabète, et l'hypertension)⁷⁹.

Ce corpus a pour objectif de définir une grammaire permettant de reconnaître de manière automatique les opérateurs déontiques dans le contexte des Guides de Bonnes Pratiques. La diversité des documents collectés permet de réunir des exemples de variabilité d'écriture de documents médicaux et d'améliorer ainsi la couverture syntaxique de reconnaissance des opérateurs déontiques dans les Guides de Bonnes Pratiques. Ce corpus a été volontairement étendu à d'autres types de documents médicaux pour augmenter la couverture syntaxique.

6.2.2 Opérateurs déontiques en contexte

Une analyse distributionnelle des verbes déontiques est réalisée à l'aide du concordancier « *Simple Concordance Program* (4.08)⁸⁰ » (partie A de la Figure 6.62 à la Figure 6.68) qui met en évidence les contextes d'occurrences des verbes déontiques à partir desquels certaines constructions particulières peuvent être décrites par un automate (partie D), par exemple *doivent contribuer à faire considérer* ou *devra souvent associer*. L'ensemble des automates (partie D) permet d'aboutir à un automate global (partie C) qui intègre l'ensemble des motifs contenant les opérateurs déontiques construits autour du verbe *devoir*.

⁷⁹ Les intitulés et les sources de ces documents sont détaillés dans l'annexe 3.

⁸⁰ Téléchargeable à partir du site : <http://www.download.com> (dernier accès le 10 juin 2006)

L'automate décrit dans la partie D de la Figure 6.62 reconnaît l'opérateur déontique *devoir* à la 3^{ème} forme du pluriel du présent de l'indicatif, suivi d'un verbe à l'infinitif (*contribuer*), d'une préposition (*à*) et de deux verbes à l'infinitif (*faire considérer*).

Le second exemple de motif est la reconnaissance de la 3^{ème} forme du singulier du verbe *devoir* au futur suivi de l'adverbe *souvent*, puis le verbe à l'infinitif *associer*. L'ensemble des variantes morphologiques des verbes analysés permet de constituer la grammaire. L'utilisation d'un étiqueteur n'est pas nécessaire dans notre cas du fait que les automates hybrides incorporent des catégories syntaxiques dans leurs nœuds terminaux.

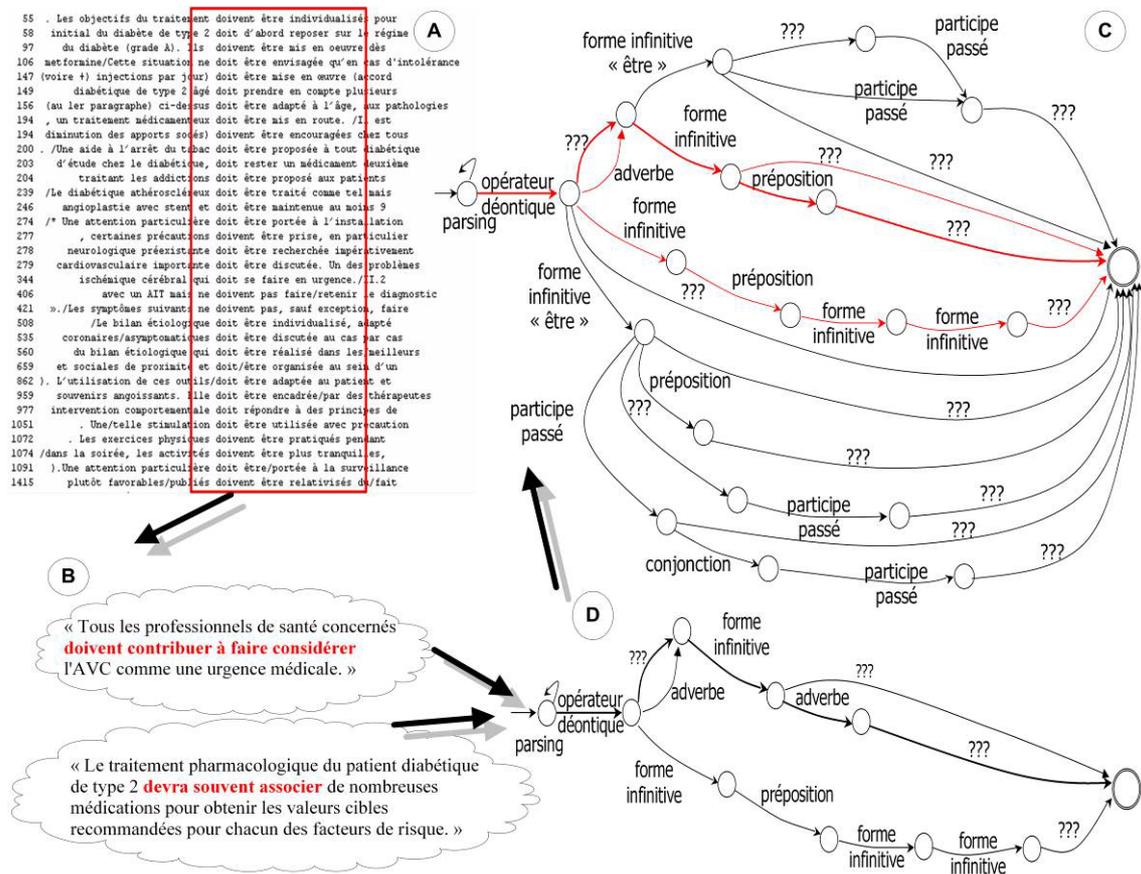


Figure 6.62 – Contextes du verbe déontique *devoir* et représentation de l'automate global correspondant.

6.2. Règles de description syntaxique

Une analyse distributionnelle du verbe déontique *pouvoir* est également réalisée (partie A de la Figure 6.63). Un des motifs (partie B) issu de cette analyse distributionnelle concerne la négation du verbe déontique à la 3^{ème} personne du singulier de l'indicatif à la voix passive (*ne peut être actuellement recommandé*). Le groupe verbal est donc caractérisé par la négation (*ne*), du verbe déontique *pouvoir* à la 3^{ème} personne du singulier de l'indicatif (*peut*), de l'auxiliaire *être*, d'un adverbe (*actuellement*) et du participe passé au singulier du verbe *recommander* (*recommandé*).

L'ensemble des automates (partie D) décrivant la négation du verbe déontique *pouvoir* permet d'aboutir à un automate général (partie C) décrivant l'ensemble des motifs construits autour du verbe *pouvoir*. Cet ensemble de variantes morphologiques permet la constitution de la grammaire pour reconnaître l'opérateur déontique caractérisé par le verbe *pouvoir*.

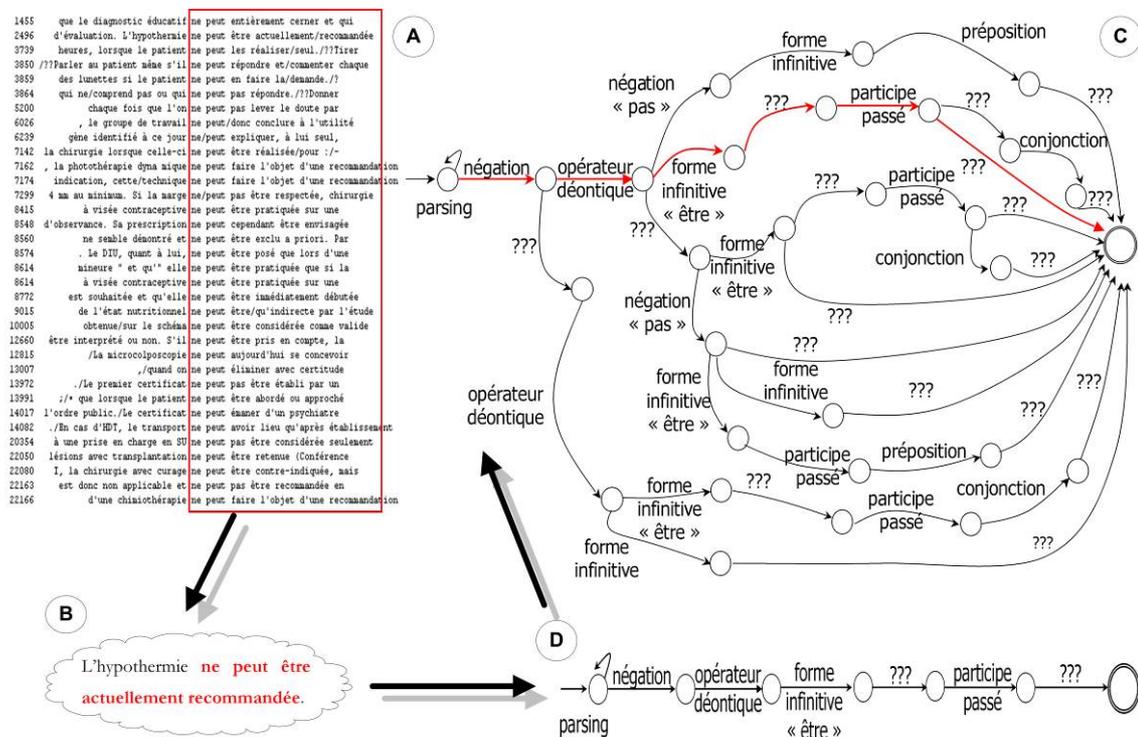


Figure 6.63 – Contextes du verbe déontique *pouvoir* à la forme négative et à la voix passive, représentation de l'automate générique correspondant.

Une analyse distributionnelle de la forme impersonnelle du verbe déontique *devoir* est réalisée (partie A de la Figure 6.64). Chaque motif syntaxique de la partie B, par exemple *Il doit rechercher*, est décrit par un automate (partie D). L'ensemble des automates (partie D) décrivant chaque motif permet d'aboutir à un automate général (partie C) qui décrit l'ensemble des motifs contenant les opérateurs déontiques construits autour de la forme impersonnelle du verbe *devoir*.

L'automate décrit dans la partie D de la Figure 6.64 reconnaît donc la forme impersonnelle de l'opérateur déontique à la 3^{ème} personne du singulier de l'indicatif (*il doit*), suivi du verbe à l'infinitif *rechercher*.

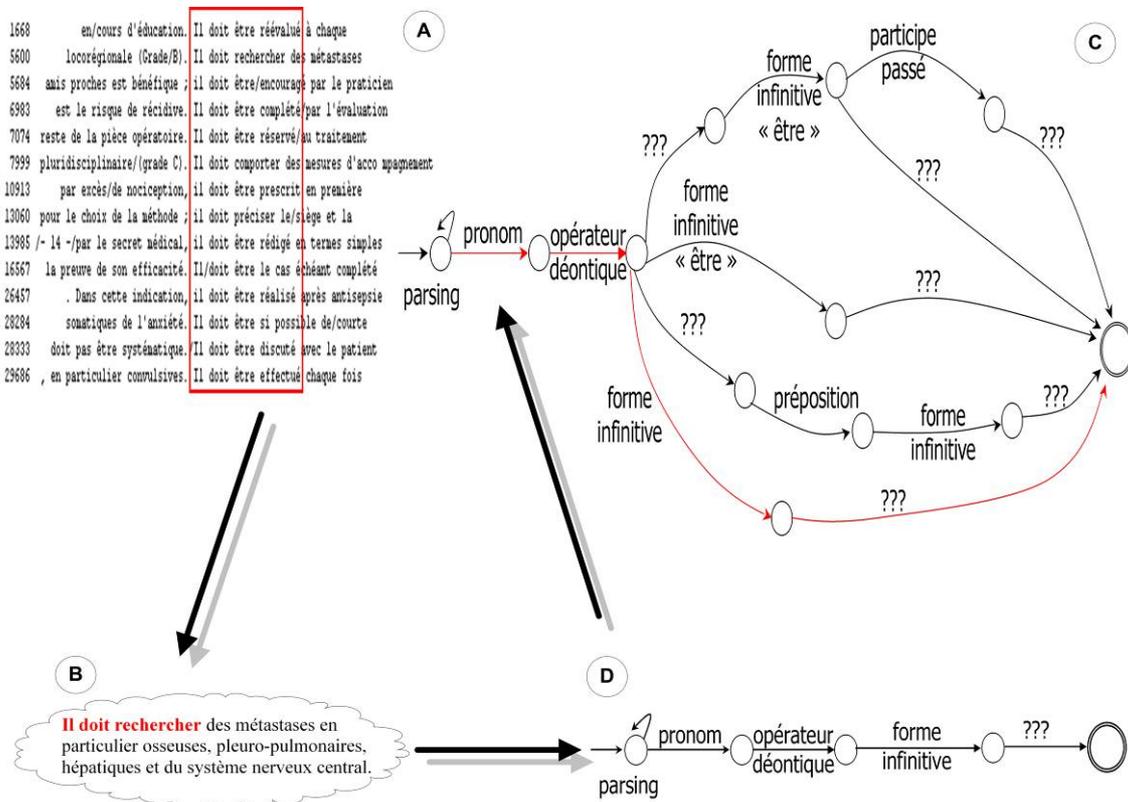


Figure 6.64 – Contextes du verbe déontique *devoir* à la forme impersonnelle, et représentation de l'automate général correspondant.

6.2. Règles de description syntaxique

Une analyse distributionnelle des verbes déontiques *recommander* et *déconseiller* est réalisée (partie A de la Figure 6.65). Chaque motif syntaxique (partie B) issu de cette analyse distributionnelle est décrit par un automate (partie D). Les deux motifs concernent les verbes déontiques à la 3^{ème} personne du singulier de l'indicatif à la voix passive (*est recommandé* et *est classiquement déconseillée*). Comme décrit précédemment, l'ensemble des automates (partie D) permet d'aboutir à un automate général (partie C) qui décrit l'ensemble des motifs contenant les opérateurs déontiques construits autour des verbes *recommander* et *conseiller*.

L'automate de la partie D de la Figure 6.65 reconnaît l'opérateur déontique *recommander* à la 3^{ème} personne du singulier de l'indicatif à la voix passive (*est recommandé*). Le second exemple représente la reconnaissance du groupe verbal caractérisé par l'auxiliaire *être* au singulier (*est*), d'un adverbe (*classiquement*) et du participe passé au singulier (*déconseillée*).

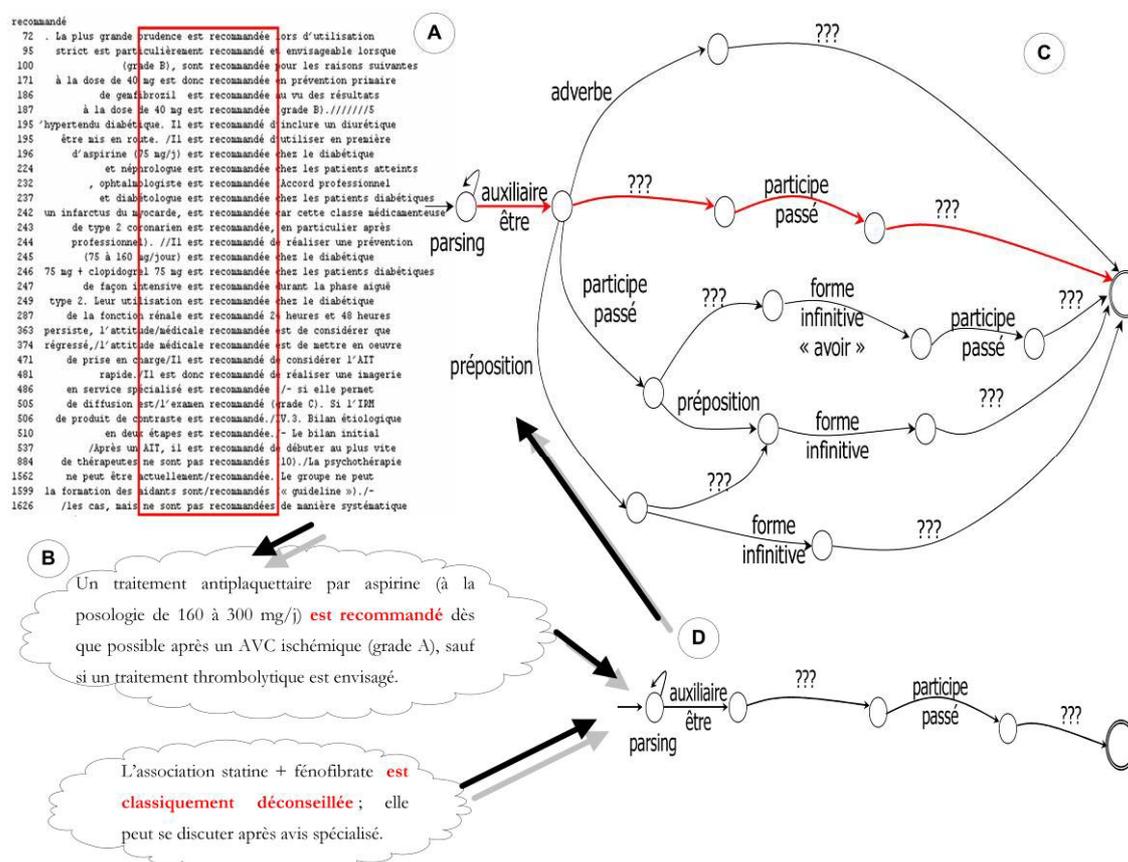


Figure 6.65 – Contextes des verbes *recommander* et *déconseiller* à la voix passive et représentation de l'automate général correspondant.

Une analyse distributionnelle concerne la forme négative du verbe déontique *recommander* (partie A de la Figure 6.66). L'automate représenté par la partie D décrit le motif syntaxique (partie B) *n'est pas recommandé*. Le groupe verbal est caractérisé par la négation (*n'*), l'auxiliaire être au singulier (*est*), de la négation (*pas*) et du participe passé au singulier (*recommandé*). La phrase est reconnue par l'automate (partie D) décrivant la négation du verbe déontique. L'ensemble des automates permet d'aboutir à un automate général (partie C) qui décrit l'ensemble des motifs contenant les négations des opérateurs déontiques construits autour du verbe *recommander*.

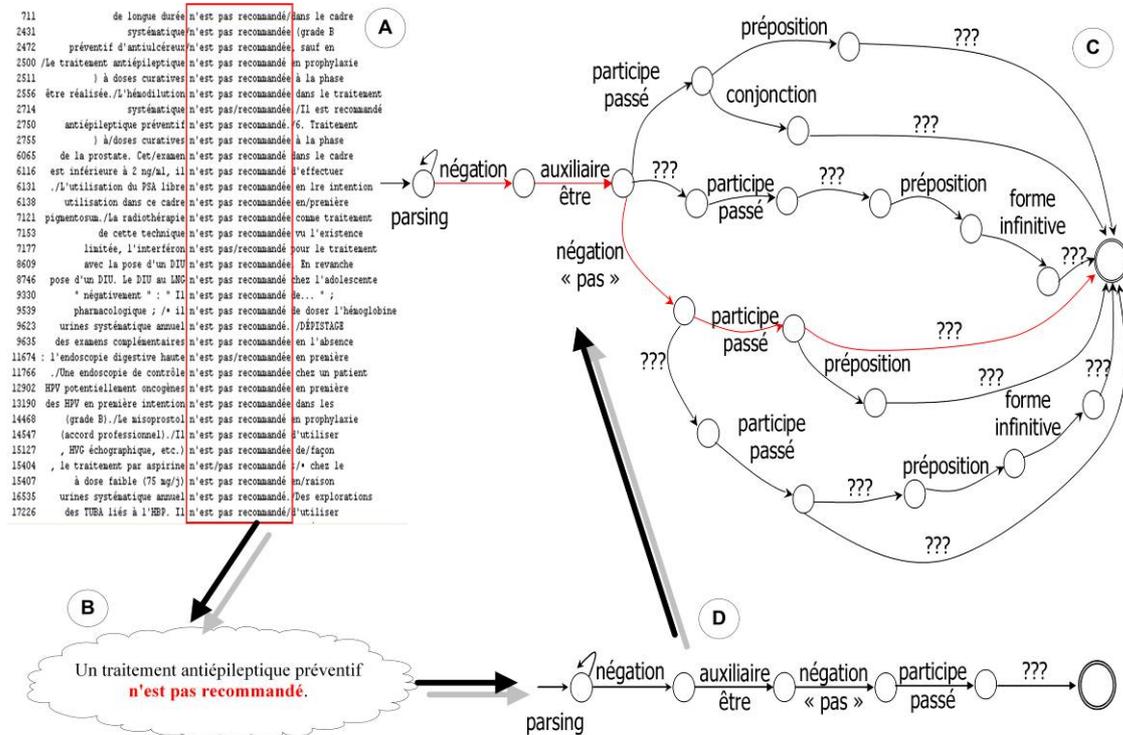


Figure 6.66 – Contextes de l'opérateur déontique *recommander* à la voix passive négative et représentation de l'automate général correspondant.

6.2. Règles de description syntaxique

Une analyse distributionnelle de la forme impersonnelle du verbe déontique *recommander* est réalisée (partie A de la Figure 6.67). Chaque motif syntaxique (partie B) issu de cette analyse distributionnelle est décrit par un automate (partie D), par exemple *il est recommandé de prescrire*. L'ensemble des automates (partie D) décrivant la forme impersonnelle du verbe déontique permet d'aboutir à un automate général (partie C) qui décrit l'ensemble des motifs contenant les opérateurs déontiques construits autour de la forme impersonnelle du verbe *recommander*.

L'automate décrit dans la partie D de la Figure 6.67 reconnaît la forme impersonnelle de l'opérateur déontique à la 3^{ème} personne du singulier de l'indicatif (*il est recommandé*), suivi de la préposition *de* et du verbe à l'infinitif *prescrire*.

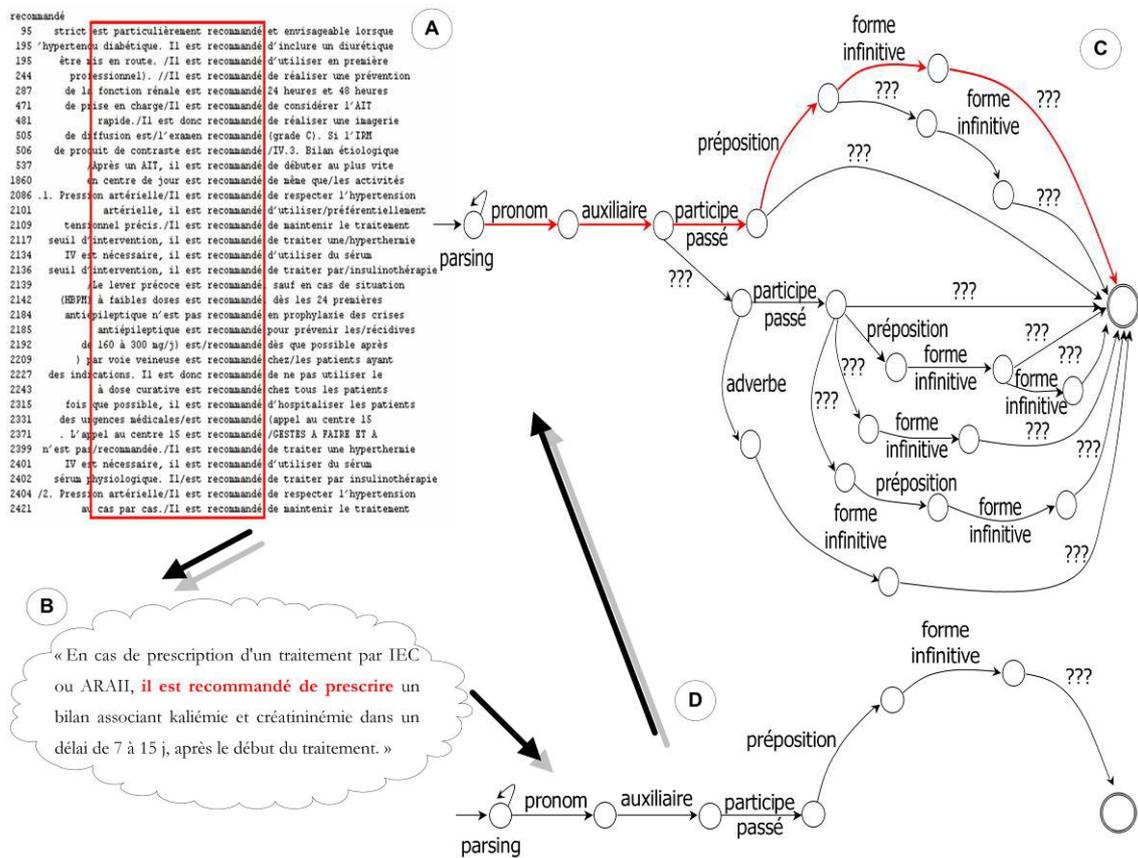


Figure 6.67 – Contextes du verbe *recommander* à la forme impersonnelle et représentation de l'automate général correspondant.

Une analyse distributionnelle de la forme impersonnelle négative du verbe déontique *recommander* est réalisée (partie A de la Figure 6.68). Le motif (partie B) issu de cette analyse distributionnelle concerne la négation du verbe déontique à la 3^{ème} personne du singulier de l'indicatif (*n'est pas recommandé*). Le groupe verbal est caractérisé par la forme impersonnelle (*il*), de la négation (*n'*), de l'auxiliaire *être* au singulier (*est*), la négation (*pas*) et du participe passé au singulier (*recommandé*) suivi d'une préposition et d'un verbe à l'infinitif (*d'utiliser*). La phrase est reconnue par l'automate (partie D) décrivant la forme impersonnelle négative du verbe déontique. L'ensemble des automates décrivant les motifs construits autour de la forme impersonnelle du verbe *recommander* permet d'aboutir à un automate général (partie C).

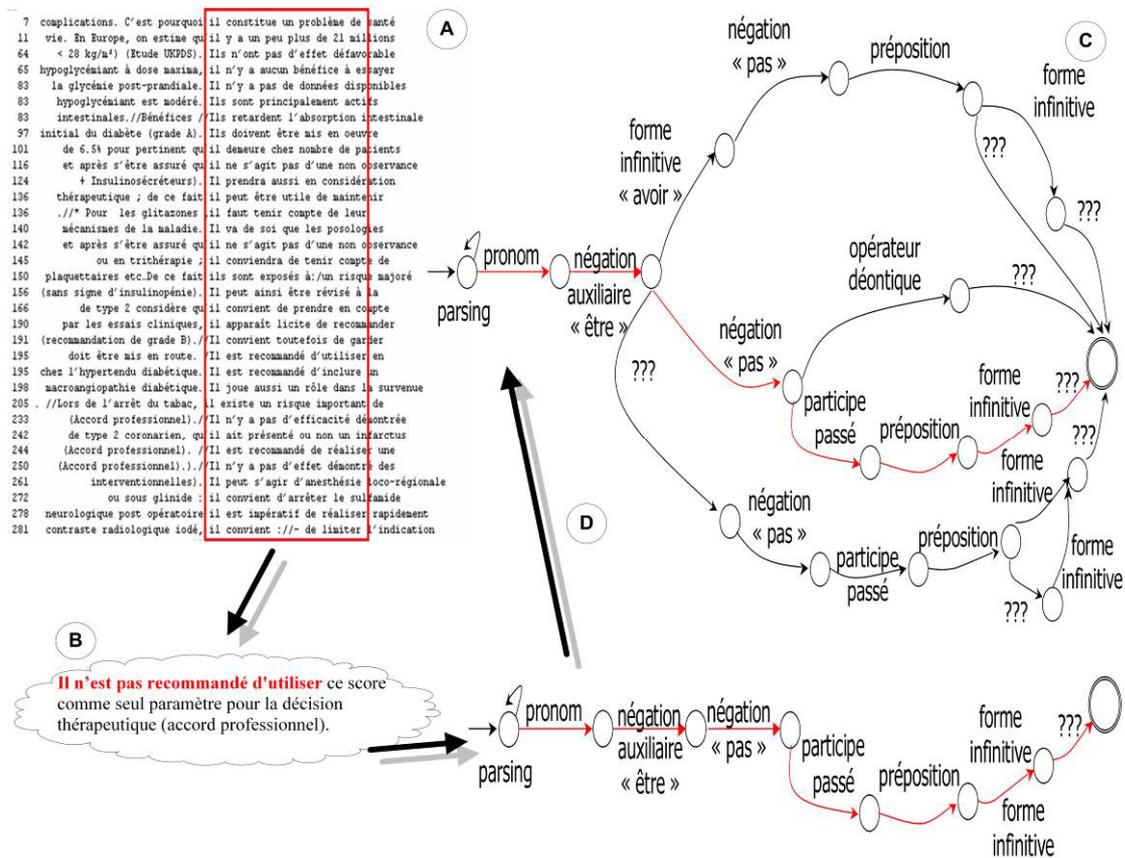


Figure 6.68 – Contextes du verbe déontique *recommander* à la forme impersonnelle négative et représentation de l'automate général correspondant.

6.2.3 Forme monadique ou dyadique des opérateurs déontiques

En se basant sur les travaux de Moulin et al. (Moulin & Rousseau, 1990), on peut définir les opérateurs déontiques monadiques suivants *il faut*, *il est obligatoire de*, *il faut que*, *il est nécessaire de-ou-que*, *il est interdit de*, *il est possible de* et les opérateurs déontiques dyadiques suivants *doit*, *doivent*, *peut*, *peuvent*, *ne doit pas*, et toutes leurs variantes morphologiques. La reconnaissance de la forme monadique *il est recommandé* est représentée par la figure ci-dessous.

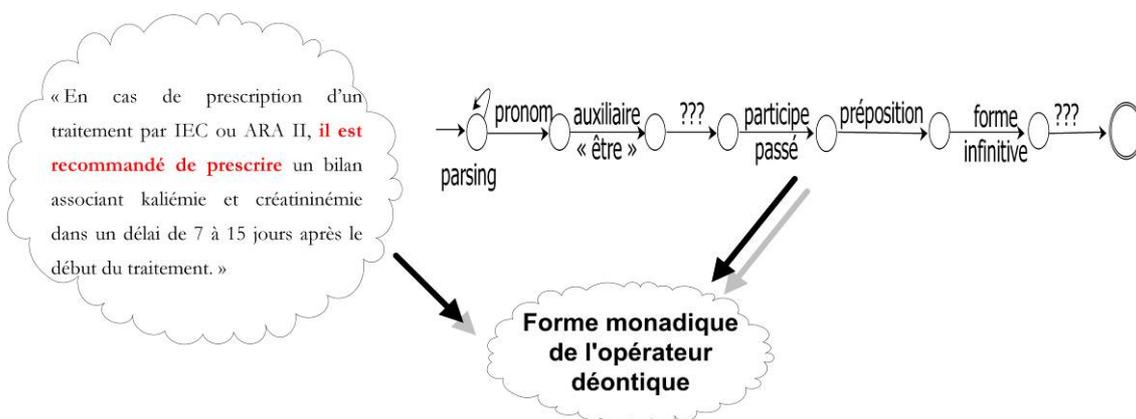


Figure 6.69 – Reconnaissance de la forme monadique de l'opérateur déontique.

Les formes monadiques et dyadiques des opérateurs sont des marqueurs pour caractériser les implications du *front-scope* et du *back-scope*⁸¹ en terme de représentation des connaissances, i.e. la partie textuelle correspondant aux prémisses d'une règle de décision, et celle correspondant aux actions.

6.2.4 Voix active ou passive des opérateurs déontiques

La voix active ou passive de l'opérateur déontique est également une caractéristique importante et nécessaire en terme de représentation des connaissances.

Les objectifs sont les mêmes que ceux concernant l'identification de la forme monadique ou dyadique de l'opérateur déontique. La voix active ou passive de l'opérateur déontique associée à la forme monadique ou dyadique permet d'établir quelques règles pour identifier quelle partie textuelle correspond aux prémisses et aux actions d'une règle de décision. La reconnaissance de la voix passive ou active des opérateurs déontiques repose sur une heuristique de reconnaissance basée sur les règles grammaticales (Figure 6.70)⁸².

⁸¹ Nous reviendrons sur cette caractérisation dans le chapitre 7 concernant la génération automatique des règles de décision ou d'éléments GEM à partir de la phrase balisée.

⁸² La voix de la phrase sera intégrée dans le processus de balisage de l'opérateur déontique permettant ainsi de dissocier les traitements automatiques à réaliser, tels que ceux intervenant pour la génération de règles de décision.

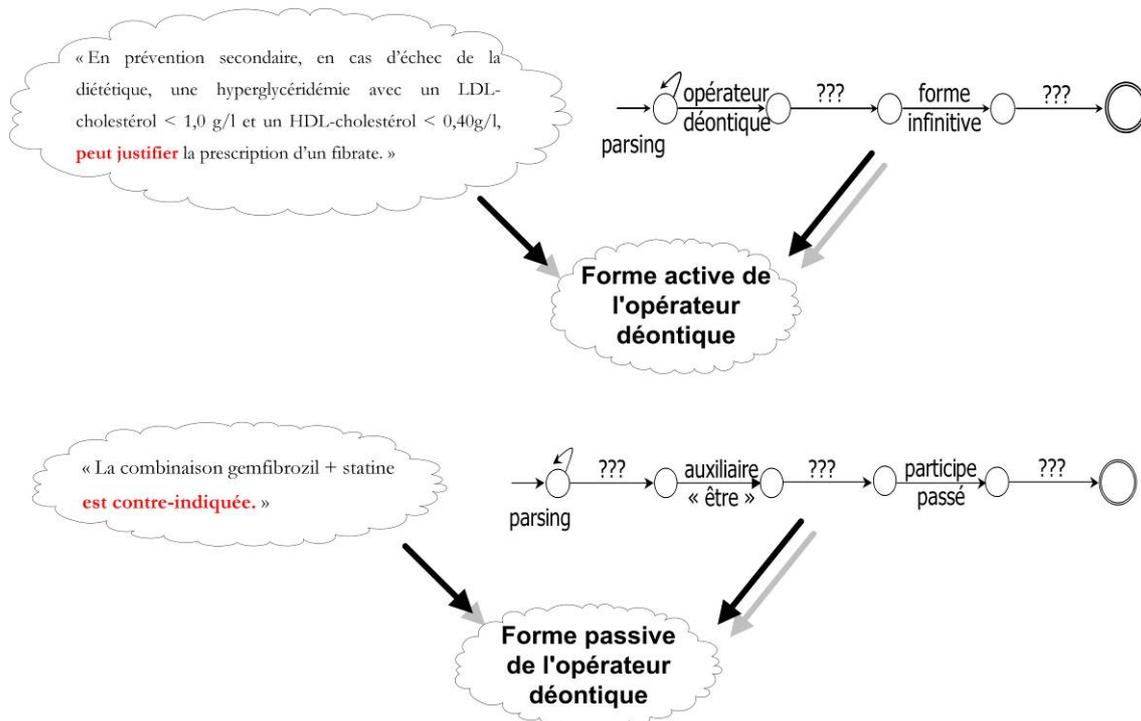


Figure 6.70 – Reconnaissance de la voix passive ou active de l'opérateur déontique.

6.2.5 Forme singulière et plurielle des opérateurs déontiques

La forme singulier/pluriel de l'opérateur déontique est représentée par les automates ci-dessous. Le premier motif concerne le verbe déontique à la 3^{ème} personne du singulier de l'indicatif à la voix passive (*est contre-indiquée*). Le groupe verbal est donc caractérisé par l'auxiliaire *être* au singulier (*est*) et du participe passé au singulier (*contre-indiquée*). Le second motif représente la reconnaissance de l'opérateur déontique au pluriel. Le groupe verbal est constitué de l'auxiliaire *être* au pluriel (*sont*) à la 3^{ème} personne du pluriel de l'indicatif à la voix passive et du participe passé au pluriel (*recommandés*).

La phrase qui ne peut pas être reconnue par les automates que nous avons décrit est une phrase contenant un auxiliaire *être* au singulier (*est*) suivi du participe passé au pluriel *recommandés*. De la même manière, le premier exemple de la figure ci-dessous n'est pas reconnu par le second automate qui reconnaît uniquement les formes plurielles des opérateurs déontiques, c'est-à-dire un groupe verbal constitué d'un auxiliaire et d'un participe passé au pluriel.

En résumé, les motifs que nous avons décrits sont représentés de telle sorte qu'on puisse identifier la forme singulier/pluriel du groupe verbal déontique, cela dans l'objectif de réduire le nombre d'opérations nécessaires à l'identification du « bon » groupe verbal.

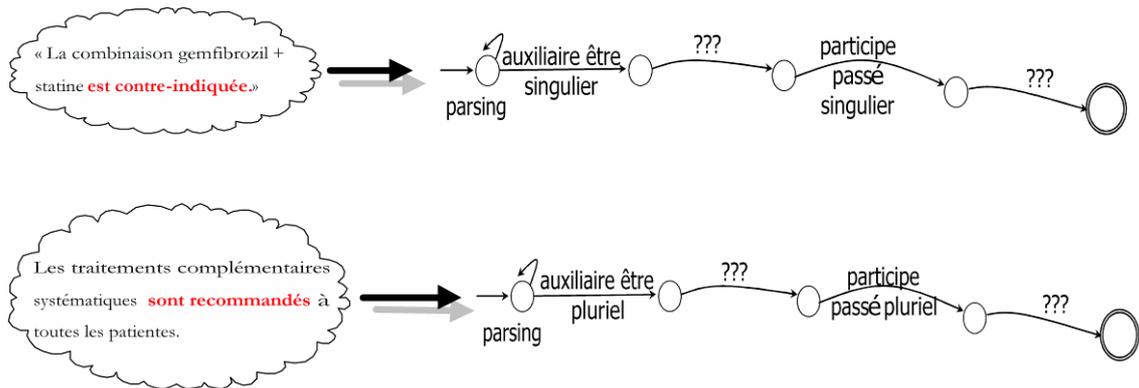


Figure 6.71 – Automates décrivant les motifs de forme singulier/pluriel.

6.2.6 Formes infinitives

Les formes infinitives des opérateurs déontiques ne posent pas de difficultés particulières et la simple reconnaissance du verbe à l’infinitif permet de le délimiter en fonction de l’opérateur déontique (Figure 6.72). La délimitation de la reconnaissance de l’opérateur déontique englobe le verbe à l’infinitif⁸³.

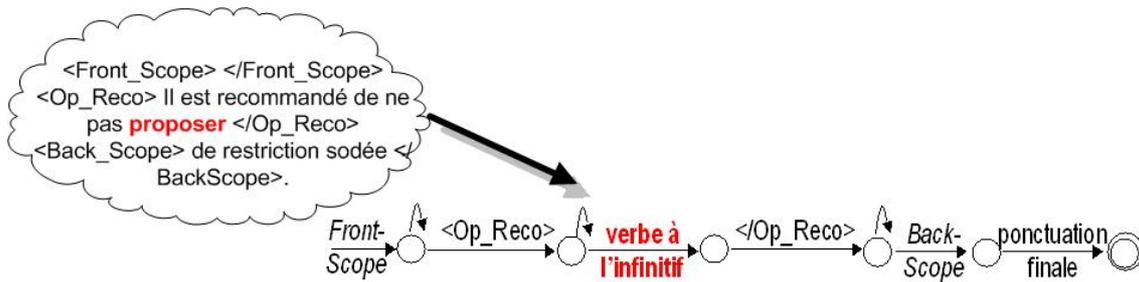


Figure 6.72 – Automate décrivant la reconnaissance des formes infinitives dans l’opérateur déontique.

⁸³ Nous expliquerons dans le chapitre suivant l’intérêt de baliser le verbe à l’infinitif. Cela est essentiellement dû au fait de proposer une règle de décision sans nuance d’interprétation.

6.2.7 Les connecteurs : marqueurs du contenu des *scopes*

Dans cette section, nous nous intéressons plus particulièrement à la reconnaissance des connecteurs, par exemple ceux exprimant des conditions. L'intérêt est principalement de pouvoir caractériser le contenu des *scopes*, et de faciliter la génération automatique d'éléments GEM ou de règles de décision à partir de la phrase balisée. La fréquence⁸⁴ des connecteurs d'un corpus d'environ 20 200 phrases de Guides de Bonnes Pratiques est représentée par la Figure 6.73.

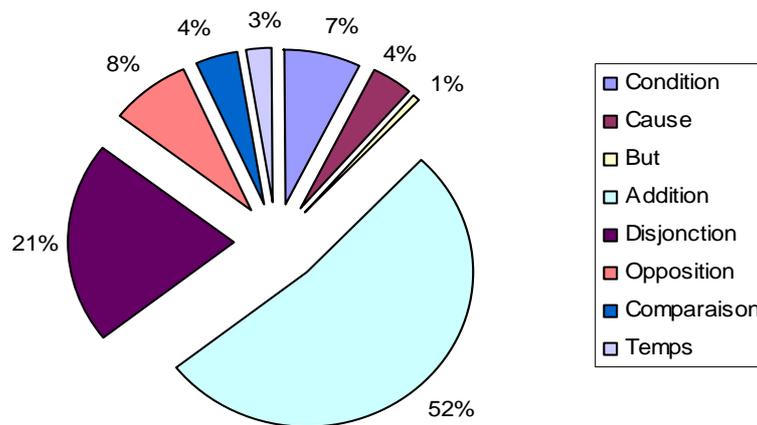


Figure 6.73 – Répartition des connecteurs dans les Guides de Bonnes Pratiques.

Les connecteurs d'addition et de disjonction ne sont pas des connecteurs significatifs à reconnaître pour générer, par exemple, de manière automatique des règles de décision ou des éléments GEM. L'objectif de la reconnaissance de certains connecteurs est principalement de caractériser des variables correspondant à des décisions ou des actions en terme de traitements de formes de surface dans les *scopes*. Les principaux connecteurs pour ce type de reconnaissance correspondent aux connecteurs de condition (7%) et de temps (3%).

Un ensemble d'automates est décrit dans la suite de ce chapitre dédié à la reconnaissance des connecteurs de condition. Ils concernent uniquement les phrases préalablement balisées⁸⁵ contenant des opérateurs déontiques.

Une analyse distributionnelle du connecteur de condition *sauf si* a été réalisée⁸⁶. Les motifs syntaxiques issus de cette analyse distributionnelle sont décrits par un automate spécifique

⁸⁴ Nous avons utilisé Tropes (<http://www.acetic.fr> (dernier accès le 10 juin 2006)) pour analyser ce corpus. Les fréquences des connecteurs sont automatiquement calculées dans ce logiciel.

⁸⁵ Une phrase non balisée ne correspond pas à une recommandation. Le balisage des connecteurs pour des phrases ne représentant pas de recommandations ne fait pas l'objet de cette recherche du fait que nous nous intéressons essentiellement dans un premier temps aux recommandations.

⁸⁶ L'analyse distributionnelle a été réalisée avec le concordancier *Simple Concordance Program*, comme cela a été le cas pour les analyses distributionnelles concernant les opérateurs déontiques. Il en sera de même pour les analyses distributionnelles réalisées pour les autres connecteurs de condition, et de temps. Nous ne détaillons pas tous les connecteurs de condition et de temps du fait qu'un sous-ensemble d'exemples permet de comprendre le processus de reconnaissance et de balisage de ces connecteurs.

6.2. Règles de description syntaxique

(Figure 6.74). L'automate décrit dans cette figure reconnaît les balises (de début et de fin) du *front-scope*, puis reconnaît les balises de l'opérateur déontique et les balises du *back-scope*. Du fait que deux opérateurs déontiques sont décomptés dans cette phrase, l'automate reconnaît à nouveau les balises d'ouverture du deuxième *front-scope*, puis le connecteur *sauf si*, suivi de la balise de fermeture du *front-scope*, et des balises du second opérateur déontique et des balises du *back-scope*. L'automate reconnaît également le segment textuel correspondant à la condition *un traitement thrombolytique*, positionné après le connecteur et avant la balise de fermeture du *front-scope*.

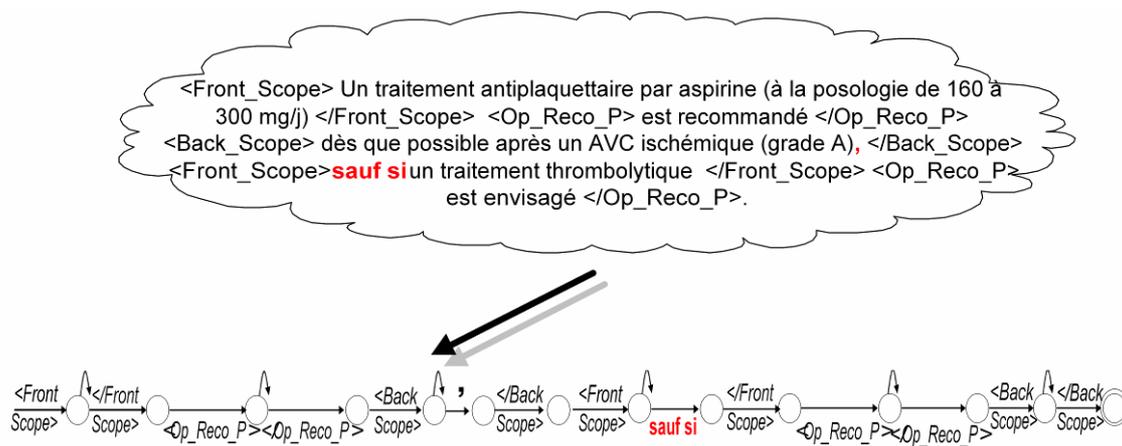


Figure 6.74 – Automate permettant la reconnaissance d'une condition par « *sauf si* ».

De la même manière, une analyse distributionnelle du connecteur de condition *si* a également été réalisée⁸⁶. Les motifs syntaxiques issus de cette analyse distributionnelle sont décrits par un automate spécifique (Figure 6.75). L'automate décrit dans cette figure reconnaît la balise (de début) du *front-scope*, du connecteur (*si*) suivi de la virgule, puis de la balise (de fermeture) du *front-scope*. Les balises de l'opérateur déontique sont reconnues ainsi que celles correspondantes au *back-scope*. L'automate reconnaît le segment textuel correspondant à la condition *le taux de CA est initialement élevée*, positionné après le connecteur et avant la virgule et de la balise (de fermeture) du *front-scope*.

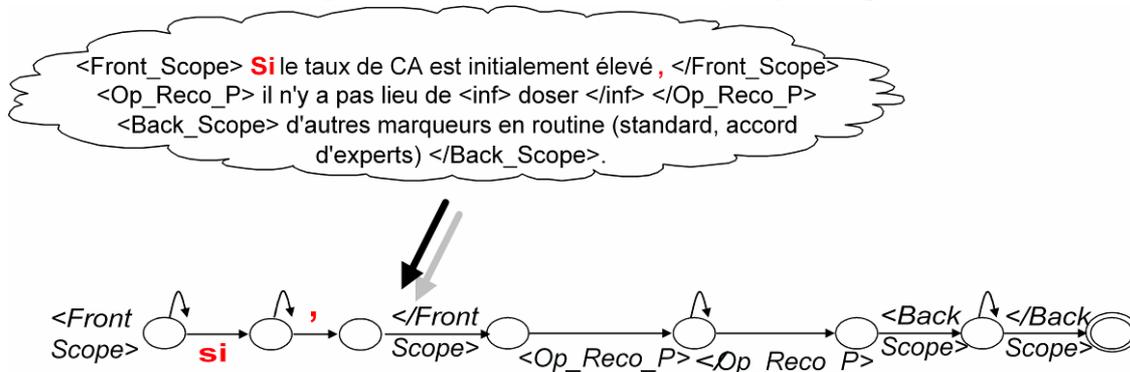


Figure 6.75 – Automate permettant la reconnaissance de la condition par « *si* » et la virgule.

En ce qui concerne le connecteur *en cas de*, les motifs syntaxiques issus de son analyse distributionnelle sont décrits par un automate spécifique (Figure 6.76). L'automate décrit dans cette figure reconnaît la balise de début du *front-scope*, la virgule et le connecteur *en cas de* suivi d'une virgule, puis de la balise de fermeture du *front-scope* ainsi que les balises de l'opérateur déontique et du *back-scope*. Comme pour toutes les reconnaissances des connecteurs, le segment textuel correspondant à la condition *intolérance ou d'efficacité insuffisante des autres statines*⁸⁷ est reconnu par l'automate.

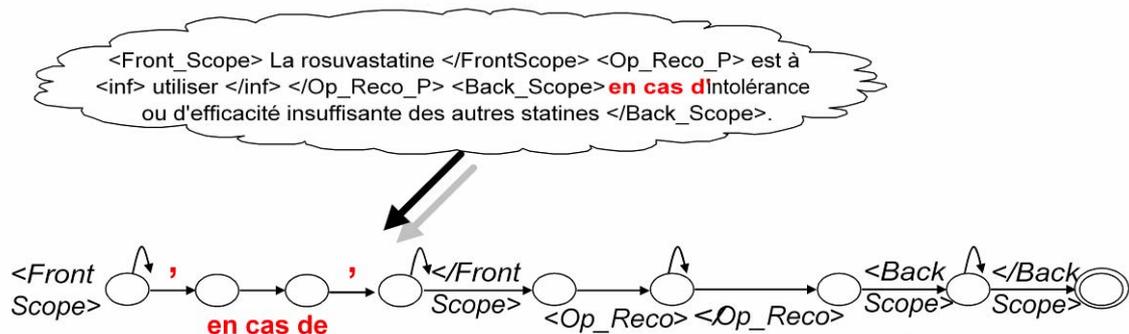


Figure 6.76 – Automate permettant la reconnaissance de la condition par « en cas de ».

Nous avons un ensemble de connecteurs de condition comme *lorsque*, *sauf en cas de*, qui s'appuie sur les mêmes principes de reconnaissance que les deux exemples exposées ci-dessous. Pour conclure cette section, la reconnaissance des conditions se base essentiellement sur la reconnaissance des balises existantes telles que le *front-scope* ou le *back-scope*, et de la ponctuation.

6.3 Stratégie de pré-traitement pour l'analyse de « textes réels »

Le pré-traitement des phrases à analyser est une étape fréquemment rencontrée dans le domaine de l'Extraction d'Informations. Cette étape est rendue nécessaire par le besoin de traiter des phrases issues de textes réels au moyen de méthodes de TAL souvent empiriques⁸⁸. La segmentation est une étape importante dans ce type de système. En effet, elle impacte directement sur la précision de l'analyse. Certains utilisent la ponctuation pour résoudre les problèmes d'anaphores (Lee et al., 2001). D'autres comme le système SNAP (Moldovan et al., 1993) effectuent une segmentation de la phrase par la reconnaissance des noms propres, du nombre de « semi-auxiliaires » comme *used to*, des groupes verbaux, des adverbes, de la ponctuation, des prépositions. Un autre exemple de pré-traitement se rencontre dans les systèmes dits de « questions réponses » (*Question Answering*) et implique

⁸⁷ Un traitement supplémentaire consisterait à identifier la particule « d' » dans la condition. Ce qui permettrait dans la génération automatique des règles de décision ou d'éléments GEM de ne pas avoir cette particule.

⁸⁸ Bien que le pré-traitement soit aussi nécessaire pour des systèmes utilisant des techniques plus sophistiquées, comme dans le système TACITUS mentionné ci-après.

L'exécution de deux tâches : (i) l'une permettant de détecter le type de réponse attendu, et (ii) l'autre de convertir une question posée en langage naturel en une liste de mots clés prioritaires à être utilisés pour la recherche dans le document ou dans le passage textuel (Ferres et al., 2005). Le pré-traitement peut être vu comme une étape qui consiste à prédire les types d'événements décrits comme cela a été le cas pour le système Paramax (Weir & Silk, 1992).

Dans le but de réaliser une analyse syntaxique plus robuste, le système TACITUS (Hobbs, Appelt & Bear, 1991) se base quant à lui sur trois techniques : (i) '*an agenda-based scheduling parser*', (ii) '*a recovery technique for failed parses*', et (iii) une nouvelle technique appelée '*terminal substring parsing*'. La dernière technique est généralement utilisée pour les phrases contenant plus de 60 mots, et permet un traitement plus rapide des phrases analysées sans en perdre la précision. La phrase est segmentée en sous-chaînes en fonction des signes de ponctuation et plus particulièrement la virgule, les conjonctions, et les pronoms relatifs. L'intérêt d'utiliser cette technique d'analyse par sous-chaînes terminales est d'analyser une petite quantité de mots, sans perdre la possibilité d'analyser toute la phrase.

En s'inspirant des pré-traitements réalisés dans les systèmes d'Extraction d'Informations, le module de pré-traitement développé dans G-DEE consiste à segmenter la phrase analysée en fonction de la ponctuation (comme décrit dans les systèmes précédents) et, de façon plus spécifique, en fonction du nombre d'occurrences des opérateurs déontiques. La segmentation de la phrase permet d'identifier les segments de textes pertinents, c'est-à-dire ceux contenant potentiellement les opérateurs déontiques.

6.3.1 Pré-traitements et Guides de Bonnes Pratiques

Le style rédactionnel relativement libre rencontré dans les Guides de Bonnes Pratiques rend nécessaire certains pré-traitements au niveau de la phrase dont le but consiste à isoler les occurrences des différents opérateurs déontiques et de leur portée respective, en particulier pour les phrases susceptibles de contenir plusieurs opérateurs déontiques. Par exemple, une même phrase peut contenir deux opérateurs déontiques, en caractères gras dans la phrase suivante, avec une ambiguïté possible quant à la portée de ces opérateurs.

« Un traitement antiplaquettaire par aspirine (à la posologie de 160 à 300 mg/j) **est recommandé** dès que possible après un AVC ischémique (grade A), sauf si un traitement thrombolytique **est envisagé**. »

La phrase ci-dessous constitue un exemple de plusieurs occurrences de l'auxiliaire être (soulignés dans le texte) dans le contexte de la reconnaissance du verbe déontique *recommander* à la forme passive. Une étape de pré-traitement peut faciliter l'identification de la forme conjuguée sans pour autant avoir à effectuer une analyse syntaxique de la phrase dans son intégralité.

« Une bithérapie orale est **conseillée** lorsque l'HbA1c est > 6,5 % sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie

*optimale, elle **est recommandée** lorsque l'HbA1c **est** > 8% , sous monothérapie orale à posologie optimale (accord professionnel). »*

Un ensemble de règles de décomposition, décrites dans la suite de ce chapitre, définit les priorités selon la ponctuation de la phrase, le nombre d'opérateurs déontiques, et le nombre d'auxiliaires.

Le module de pré-traitement, qui repose sur un découpage propositionnel du texte, est constitué de trois modules. Le premier consiste à décompter respectivement : (1) le nombre d'occurrences de certaines catégories de mots tels que les groupes verbaux ; (2) le nombre d'opérateurs déontiques ; (3) la ponctuation et les conjonctions de coordination. Le deuxième module consiste à identifier les phrases susceptibles d'être segmentées, et le troisième consiste à segmenter la phrase pour isoler les occurrences des opérateurs déontiques. Le pré-traitement se décompose selon : i) un aspect heuristique basé sur les décomptes des verbes déontiques, dont les ratios vont orienter vers des stratégies de traitement déterminées et ii) un aspect pratique, à base de segmentation, qui sert directement d'entrée aux modules d'analyse.

6.3.2 Nécessité d'un pré-traitement

Lorsque la phrase considérée contient un seul opérateur déontique, il n'est pas nécessaire de segmenter la phrase. La reconnaissance d'un opérateur déontique unique repose sur l'identification d'un seul motif syntaxique et d'une portée sans ambiguïté. La phrase ci-dessous illustre l'exemple d'une telle configuration où un seul opérateur déontique *devoir* (en caractères gras dans l'exemple) à la forme passive a été détecté.

*« Un ECG **doit être réalisé** dès le début de la prise en charge. »*

De la même manière, aucun pré-traitement n'est rendu nécessaire pour l'identification de la forme conjuguée de l'opérateur déontique *recommander* (en caractères gras dans l'exemple) dans la phrase suivante. Un seul auxiliaire *être* est décompté, ce qui n'implique pas d'ambiguïté avec la forme conjuguée de l'opérateur déontique.

*« En cas de prescription d'un traitement par IEC ou ARAII, **il est recommandé de prescrire** un bilan associant kaliémie et créatininémie dans un délai de 7 à 15 jours, après le début du traitement. »*

6.3.3 Règles d'usage de la ponctuation

Les règles de la ponctuation (Drillon, 1991) permettent d'établir un ensemble de règles spécifiques de pré-traitement. L'exemple ci-dessous représente une phrase contenant deux opérateurs déontiques (en caractères gras dans l'exemple), correspondant à la forme passive du verbe *pouvoir* et à la forme active du verbe *devoir*. La virgule (en caractère gras dans l'exemple) est positionnée après le premier opérateur déontique '*ne peut être contre-indiquée*'.

6.3. Stratégie de pré-traitement pour l'analyse de « textes réels »

« En cas de ganglion pédiculaire envahi, si la résecabilité est de classe I, la chirurgie avec curage **ne peut être contre-indiquée**, mais cette décision **doit néanmoins s'intégrer** dans une approche multidisciplinaire. »

La segmentation à partir de la virgule permet de dissocier deux propositions contenant chacune un opérateur déontique. La première proposition est :

- (1) « En cas de ganglion pédiculaire envahi, si la résecabilité est de classe I, la chirurgie avec curage **ne peut être contre-indiquée**, »

et la seconde,

- (2) « mais cette décision **doit néanmoins s'intégrer** dans une approche multidisciplinaire. »

Néanmoins, certaines phrases ne peuvent être segmentées uniquement en fonction de la ponctuation. Par exemple, la phrase ci-dessous contient deux opérateurs déontiques non séparés par un signe de ponctuation. Les opérateurs déontiques correspondent au verbe *devoir*, l'un à la voix active et l'autre à la voix passive (en caractères gras dans l'exemple).

« Chez les patients ayant initialement une concentration très élevée de LDL-cholestérol, et notamment chez les patients à haut risque dont la cible thérapeutique est basse (< 1 g/l), le prescripteur **doit garder à l'esprit** que la prescription de statine à fortes doses ou en association nécessite une prise en compte au cas par cas du rapport bénéfice / risque **et ne doit jamais être** systématique. »

Cette configuration implique la recherche d'un autre marqueur de segmentation, par exemple la conjonction de coordination. Dans l'exemple ci-dessus, le marqueur de segmentation est « et » ce qui permet de segmenter la phrase en deux propositions isolant ainsi chaque opérateur déontique. La première proposition correspond donc à :

- (1) « Chez les patients ayant initialement une concentration très élevée de LDL-cholestérol, et notamment chez les patients à haut risque dont la cible thérapeutique est basse (< 1 g/l), le prescripteur **doit garder à l'esprit** que la prescription de statine à fortes doses ou en association nécessite une prise en compte au cas par cas du rapport bénéfice / risque **et** »

Et la seconde à :

- (2) « **ne doit jamais être** systématique. »

On peut s'interroger sur la signification réelle de ce type d'expression et si elle devrait être évitée dans le cadre de guides d'écriture. Toujours est-il qu'en l'état actuel, il ne semble pas fidèle à l'esprit du rédacteur de ne retenir qu'un seul des deux opérateurs déontiques.

Cette section du chapitre est dédiée aux règles de la ponctuation qui peuvent être utilisées pour définir les pré-traitements. Le signe de ponctuation le plus fréquent est la virgule.

Comme le montrent les travaux de J. McClelland (Drillon, 1991), basé sur les ouvrages du XVI^e siècle, elle peut être de différentes natures telles que :

- elliptique (« elle symbolise des éléments non répétés ») ;
- formelle (« elle marque la conjonction et la disjonction ») ;
- grammaticale (« elle éclaire le rapport syntaxique entre deux syntagmes ») ;
- particularisante (« elle équivaut à nos guillemets actuels, éventuellement à l'italique : [elle] marque les incises et isole dans le discours les signifiants dont le signifié n'est pas un mot mais un concept numérique ou un son musical ») ;
- mimétique (« elle indexe une pause respiratoire, prosodique ou rhétorique »).

Dans le but d'isoler les opérateurs déontiques en fonction de marqueurs spécifiques, seul un sous-ensemble des règles d'usage proposées est effectivement nécessaire à notre pré-traitement.

1^{ère} règle : la virgule a pour fonction de séparer plusieurs verbes sans conjonction de coordination dans la phrase.

Le marqueur de segmentation pour une phrase telle que celle présentée ci-dessous est le signe de ponctuation, et plus précisément la virgule (en caractère gras dans l'exemple ci-dessous) :

« Une bithérapie orale **est conseillée** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

Le premier opérateur déontique de la phrase correspond au verbe *conseiller* à la voix passive et la virgule le sépare du deuxième opérateur déontique (*recommander*) à la voix active. Ce marqueur de segmentation permet d'aboutir à deux propositions contenant chacune un seul opérateur déontique. La première proposition est donc :

- (1) « Une bithérapie orale **est conseillée** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, »

et la seconde :

- (2) « elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

Cette règle permet donc de segmenter la phrase en fonction de la virgule et de son positionnement par rapport aux opérateurs déontiques. En ce qui concerne le pré-traitement, nous avons défini deux modules principaux :

- (1) Décomptage préliminaire du nombre d'occurrences de virgules
- (2) Positionnement de la virgule par rapport aux opérateurs déontiques
 - a. Segmentation en sous-chaînes de la phrase si le premier opérateur déontique est reconnu.

6.3. Stratégie de pré-traitement pour l'analyse de « textes réels »

- b. Segmentation en sous-chaînes de la phrase restante si celle-ci contient la virgule.
- c. Segmentation en sous-chaînes de la phrase restante si le second opérateur déontique est reconnu dans celle-ci.

En considérant le premier exemple énoncé précédemment,

« Une bithérapie orale **est conseillée** lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

le premier module nous renvoie une seule occurrence de la virgule⁸⁹. Le second module permet de positionner cette virgule par rapport aux deux opérateurs déontiques présents dans la phrase. Pour cela, la première étape concerne la segmentation en sous-chaînes (a) et on obtient :

« lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

Puis par (b) on a :

« elle **est recommandée** lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

Et (c) nous donne :

« lorsque l'HbA1c est > 8% sous monothérapie orale à posologie optimale (accord professionnel). »

Nous pouvons alors conclure que la virgule est bien positionnée entre les deux opérateurs déontiques *conseiller* et *recommander*, et peut être considérée comme marqueur pour segmenter la phrase en deux propositions isolant chaque opérateur déontique.

2^{ème} règle : les conjonctions de coordination ont pour fonction de séparer plusieurs verbes.

Les deux opérateurs déontiques identifiés dans les exemples ci-dessous ne sont pas séparés par des signes de ponctuation. On peut néanmoins utiliser les conjonctions de coordination, plus particulièrement celles marquant les oppositions, du fait qu'elles ont précisément pour fonction de séparer les verbes. Dans l'exemple ci-dessous, nous

⁸⁹ Le nombre « 6,5% » est considéré comme une unité lexicale à part entière lors de l'analyse de la phrase. Le système identifie uniquement la virgule en tant que signe de ponctuation, c'est-à-dire définit par la règle dactylographique où un espace suit obligatoirement la virgule mais aucun espace ne précède la virgule. Ici nous ne traitons pas des problèmes dactylographiques des textes analysés qui pourraient néanmoins être préalablement vérifiés en utilisant des outils de TALN dédiés, mais n'est pas l'objet de cette thèse.

considérons donc que le marqueur de segmentation pour isoler chaque opérateur déontique est la conjonction de coordination *mais*.

Exemple 1 : « *Si le diabète est diagnostiqué chez un patient âgé, un objectif de HbA1c comprise entre 6,5% et 8,5% **peut servir de référence** mais **il est essentiel d'individualiser** cet objectif en fonction du contexte médical et social (accord professionnel).* »

Exemple 2 : « *Des explorations complémentaires des gros vaisseaux, à commencer par l'exploration écho-doppler, **ne sont pas recommandées** à titre systématique mais **seront proposées** en cas d'anomalies à l'examen clinique.* »

Pour le premier exemple, les deux propositions issues de la segmentation par le marqueur de coordination sont donc :

- (1) « *Si le diabète est diagnostiqué chez un patient âgé, un objectif de HbA1c comprise entre 6,5% et 8,5% **peut servir de référence** mais* »
- (2) « ***il est essentiel d'individualiser** cet objectif en fonction du contexte médical et social (accord professionnel).* »

Chaque proposition issue de la segmentation contient un seul opérateur déontique. Pour le second exemple, les deux propositions sont :

- (1) « *Des explorations complémentaires des gros vaisseaux, à commencer par l'exploration écho-doppler, **ne sont pas recommandées** à titre systématique mais* »
- (2) « ***seront proposées** en cas d'anomalies à l'examen clinique.* »

Pour cet exemple aussi, les opérateurs déontiques ont été correctement isolés. Nous avons également défini les deux modules permettant la segmentation de la phrase reposant sur le même processus que celui adopté pour la virgule en considérant néanmoins dans ce cas la conjonction de coordination comme le marqueur de segmentation⁹⁰ :

- (1) Décomptage du nombre d'occurrences de conjonctions de coordination (*et, mais*)
- (2) Positionnement des conjonctions par rapport aux opérateurs déontiques
 - a. Segmentation en sous-chaînes de la phrase si le premier opérateur déontique est reconnu.
 - b. Segmentation en sous-chaînes de la phrase restante si celle-ci contient la conjonction.
 - c. Segmentation en sous-chaînes de la phrase restante si le second opérateur déontique est reconnu dans celle-ci.

À la suite de ce processus, la phrase analysée contient effectivement une conjonction permettant d'isoler les opérateurs déontiques.

⁹⁰ Le module identifie au préalable qu'il n'y a pas de signe de ponctuation permettant de segmenter la phrase pour en isoler les opérateurs déontiques.

Les autres signes de ponctuation

Dans cette section, nous nous intéressons plus particulièrement au deux-points en tant que marqueur de segmentation permettant d'isoler les opérateurs déontiques, comme cela peut être le cas pour la phrase ci-dessous.

« Le cancer du sein traité **doit être considéré** comme une situation particulièrement à risque **;** **il est recommandé de rechercher** systématiquement les symptômes de détresse psychologique, anxiété, troubles sexuels, états de stress post-traumatique, et dépression, du fait de leur prévalence plus élevée et des difficultés particulières du diagnostic dans ce cas (grade B). »

Les deux propositions issues du pré-traitement sont donc :

- (1) « Le cancer du sein traité **doit être considéré** comme une situation particulièrement à risque **;** »
- (2) « **il est recommandé de rechercher** systématiquement les symptômes de détresse psychologique, anxiété, troubles sexuels, états de stress post-traumatique, et dépression, du fait de leur prévalence plus élevée et des difficultés particulières du diagnostic dans ce cas (grade B). »

Selon Grévisse, le point-virgule ne sépare pas strictement les parties d'une phrase : « en l'employant l'auteur montre au contraire la volonté délibérée de les relier, d'en montrer la nature commune ou indissociable. » Le point-virgule joue donc également le rôle d'une accolade pouvant faire s'enchaîner des propositions incomplètes. Il relie des propositions indépendantes mais associées dans une seule action. L'exemple ci-dessous illustre ce genre de configuration où le point-virgule (en caractère gras dans le texte) est considéré dans le pré-traitement comme un marqueur de segmentation.

« L'examen visuel et palpatoire des aires ganglionnaires pédiculaires et coeliaques **doit être réalisé** avant tout geste hépatique **;** tout ganglion suspect (par sa taille ou sa texture) **doit être prélevé** pour examen extemporané, l'intérêt du picking systématique n'est pas démontré. »

Les deux propositions issues de la segmentation sont :

- (1) « L'examen visuel et palpatoire des aires ganglionnaires pédiculaires et coeliaques **doit être réalisé** avant tout geste hépatique **;** »
- (2) « tout ganglion suspect (par sa taille ou sa texture) **doit être prélevé** pour examen extemporané, l'intérêt du picking systématique n'est pas démontré. »

Nous gardons la même démarche pour identifier le rôle de la ponctuation dans la phrase en la segmentant selon les étapes suivantes :

- (1) Décomptage du nombre d'occurrences de point-virgule
- (2) Positionnement du point-virgule par rapport aux opérateurs déontiques

- a. Segmentation en sous-chaînes de la phrase si le premier opérateur déontique est reconnu.
- b. Segmentation en sous-chaînes de la phrase restante si celle-ci contient le point-virgule.
- c. Segmentation en sous-chaînes de la phrase restante si le second opérateur déontique est reconnu dans celle-ci.

L'exemple ci-dessous illustre les énumérations introduites par les deux-points, et contenant un seul opérateur déontique.

« Les symptômes suivants **ne doivent pas, sauf exception, faire évoquer un AIT** ⚡ :

- altération de la conscience sans autre signe d'atteinte de la circulation vertébrobasilaire ;
- confusion mentale isolée ;
- amnésie isolée ;
- faiblesse généralisée ;
- lipothymie ;
- scotome scintillant ;
- acouphènes isolés ;
- incontinence sphinctérienne urinaire et/ou anale ;
- progression des symptômes (notamment sensitifs) selon une marche intéressant plusieurs parties du corps ;
- trouble aigu du comportement. »

Les deux propositions issues du pré-traitement sont donc :

- (1) « Les symptômes suivants **ne doivent pas, sauf exception, faire évoquer un AIT** ⚡ »
- (2) « - altération de la conscience sans autre signe d'atteinte de la circulation vertébrobasilaire ; - confusion mentale isolée ; - amnésie isolée ; - faiblesse généralisée ; - lipothymie ; - scotome scintillant ; - acouphènes isolés ; - incontinence sphinctérienne urinaire et/ou anale ; - progression des symptômes (notamment sensitifs) selon une marche intéressant plusieurs parties du corps ; - trouble aigu du comportement. »

En conservant toujours le même principe de positionnement, les deux modules sont :

- (1) Décomptage du nombre d'occurrences du deux-points
- (2) Positionnement du deux-points par rapport aux opérateurs déontiques
 - a. Segmentation en sous-chaînes de la phrase si le premier opérateur déontique est reconnu.
 - b. Segmentation en sous-chaînes de la phrase restante si celle-ci contient le deux-points.
 - c. Segmentation en sous-chaînes de la phrase restante si le second opérateur déontique est reconnu dans celle-ci.

6.3.4 Pré-traitements des phrases contenant des signes de ponctuation

Nous avons introduit cette étape de pré-traitement du fait de certaines tournures syntaxiques pouvant conduire à une mauvaise reconnaissance des opérateurs déontiques de la phrase. Ce processus nous permet d'analyser des phrases telles que celles contenant plusieurs occurrences des opérateurs déontiques et/ou celles contenant plusieurs occurrences de l'auxiliaire *être*.

Configuration 1 : La phrase contient un seul opérateur déontique à la forme passive et plusieurs occurrences de l'auxiliaire *être* ; la virgule est positionnée avant l'opérateur déontique.

La phrase contient un nombre d'occurrences de l'auxiliaire *être* supérieur au nombre d'occurrences des opérateurs déontiques. Les deux exemples ci-dessous illustrent ces situations, où les auxiliaires sont soulignés et l'opérateur déontique en caractères gras.

Exemple 1 : « Si le risque est faible ou moyen, **il est recommandé de débiter** un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque. »

Exemple 2 : « Lorsque l'HbA1c est > 8% sur deux contrôles successifs, une modification du traitement **est recommandée** (accord professionnel). »

Le marqueur de segmentation des phrases ci-dessus est la virgule (en caractère gras dans le texte), positionnée avant l'opérateur déontique *recommander*. Les deux propositions issues de la segmentation concernant l'exemple 1 sont :

- (1) « Si le risque est faible ou moyen, »
- (2) « **il est recommandé de débiter** un traitement médicamenteux en cas de persistance de chiffres élevés après traitement non médicamenteux seul et prise en charge des éventuels facteurs de risque. »

De la même manière, les propositions issues de la segmentation du second exemple sont :

- (1) « Lorsque l'HbA1c est > 8% sur deux contrôles successifs, »
- (2) « une modification du traitement **est recommandée** (accord professionnel). »

Cette segmentation, considérant la virgule comme marqueur de coupure, permet d'identifier correctement le groupe verbal, plus particulièrement le groupe verbal à la voix passive du verbe « *recommander* ».

Configuration 2 : La phrase contient deux opérateurs déontiques à la voix passive avec un nombre d'occurrences de l'auxiliaire être identique au nombre de verbes déontiques au participe passé ; la virgule est positionnée après le premier opérateur déontique.

L'exemple ci-dessous illustre l'exemple d'une phrase contenant deux opérateurs déontiques à la voix passive. Le module de pré-traitement considère la virgule (en caractère gras dans le texte) comme marqueur de segmentation, du fait qu'elle sépare les deux opérateurs déontiques.

« Un traitement antiplaquettaire par aspirine (à la posologie de 160 à 300 mg/j) ***est recommandé*** dès que possible après un AVC ischémique (grade A), ***sauf si un traitement thrombolytique est envisagé.*** »

Les deux propositions issues de la segmentation sont :

- (1) « Un traitement antiplaquettaire par aspirine (à la posologie de 160 à 300 mg/j) ***est recommandé*** dès que possible après un AVC ischémique (grade A), »
- (2) « ***sauf si un traitement thrombolytique est envisagé.*** »

Le processus reste à l'identique lorsque la phrase contient des verbes à la voix passive à des genres différents (singulier / pluriel), par exemple :

« L'utilisation d'échelles du type suivant ***est recommandée*** pour l'état neurologique, plusieurs échelles spécifiques à l'AVC ***sont utilisées***, en langue française ou anglaise ou en traduction française : échelle d'Orgogozo, échelle canadienne, échelle scandinave, échelle du National Institutes of Health (NIH). »

Le marqueur de segmentation est la virgule positionnée après le verbe *recommander*. À l'issue de cette segmentation, les deux propositions sont :

- (1) « L'utilisation d'échelles du type suivant ***est recommandée*** pour l'état neurologique, »
- (2) « ***plusieurs échelles spécifiques à l'AVC sont utilisées***, en langue française ou anglaise ou en traduction française : échelle d'Orgogozo, échelle canadienne, échelle scandinave, échelle du National Institutes of Health (NIH). »

Configuration 3 : La phrase contient deux opérateurs déontiques à la voix passive et contient plus d'occurrences de l'auxiliaire être que de verbes déontiques au participe passé ; la virgule est positionnée après le premier opérateur déontique.

La virgule est un des moyens pour séparer deux propositions juxtaposées. Nous la considérons comme un marqueur de segmentation de la phrase. Ci-dessous l'exemple d'une phrase contenant plusieurs occurrences de l'auxiliaire *être* et deux opérateurs déontiques à la voix passive, l'un étant le verbe *conseiller* et l'autre le verbe *recommander*.

6.3. Stratégie de pré-traitement pour l'analyse de « textes réels »

« Une bithérapie orale **est conseillée** lorsque l'HbA1c **est** > 6,5 % sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle **est recommandée** lorsque l'HbA1c **est** > 8% , sous monothérapie orale à posologie optimale (accord professionnel). »

La virgule étant positionnée après le premier opérateur déontique *est conseillée*, elle est le marqueur de segmentation de la phrase. Les deux propositions issues de la segmentation sont donc :

- (1) « Une bithérapie orale **est conseillée** lorsque l'HbA1c est > 6,5 % sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, »
- (2) « elle **est recommandée** lorsque l'HbA1c est > 8%, sous monothérapie orale à posologie optimale (accord professionnel). »

Les deux propositions contiennent chacune un seul opérateur déontique, et chacune contient deux occurrences de l'auxiliaire être. Un module de l'analyseur utilisant un système de « fenêtrage » permettra de reconnaître le « bon » groupe verbal. Ce module est expliqué plus précisément dans le chapitre 7.

Configuration 4 : La phrase contient deux opérateurs déontiques – un à la voix active et l'autre à la voix passive ; la virgule est positionnée après le premier opérateur déontique.

Deux opérateurs déontiques sont décomptés dans l'exemple ci-dessous, l'un à la voix active et le second à la voix passive. Le marqueur de segmentation correspond à la virgule positionnée après l'opérateur déontique *indiquer*.

« Si une restriction hydrique **est indiquée**, elle **doit être modérée**, de l'ordre de 1 000 ml/j. »

Les deux propositions sont donc :

- (1) « Si une restriction hydrique **est indiquée**, »
- (2) « elle **doit être modérée**, de l'ordre de 1 000 ml/j. »

Les deux propositions contiennent un seul opérateur déontique, et elles ont été correctement segmentées pour être analysées.

Configuration 5 : La phrase contient deux opérateurs déontiques et plusieurs occurrences de l'auxiliaire être ; la virgule est positionnée avant l'opérateur déontique.

Le module de pré-traitement décompte plusieurs virgules dans la phrase ci-dessous. Le marqueur de segmentation identifié par le pré-traitement correspond à la virgule positionnée après le verbe *recommander*.

« Lorsque l'HbA1c est > 8%, **il est recommandé de prescrire** de la metformine, et des mesures hygiéno-diététiques **doivent être associées.** »

La segmentation de la phrase permet d'identifier deux propositions :

- (1) « Lorsque l'HbA1c est > 8%, **il est recommandé de prescrire** de la metformine,
- (2) « et des mesures hygiéno-diététiques **doivent être associées.** »

La première proposition contient deux auxiliaires *être* et le module d'analyse permettra de constituer le « *bon* » groupe verbal en utilisant un système de « *fenêtrage* » (cf. Chapitre 7).

Configuration 6 : La phrase contient un opérateur déontique en fin de phrase et contient plusieurs auxiliaires.

La virgule est considérée dans l'exemple ci-dessous comme une incidente déterminative du fait que l'opérateur déontique est positionné en fin de phrase.

« Lorsque la valeur d'HbA1c est comprise entre 6,5 et 8 % une modification du traitement **peut être envisagée** en fonction de l'appréciation par le clinicien du rapport avantages/inconvénients du changement de traitement envisagé, lorsque l'HbA1c est > 8 % **elle est recommandée.** »

Dans un premier temps, le pré-traitement segmente cette phrase en fonction de la ponctuation. Ainsi, la phrase est dissociée en deux propositions à partir de la virgule. La première proposition est donc :

« Lorsque la valeur d'HbA1c est comprise entre 6,5 et 8 % une modification du traitement **peut être envisagée** en fonction de l'appréciation par le clinicien du rapport avantages/inconvénients du changement de traitement envisagé, »

Et la seconde est :

« lorsque l'HbA1c est > 8 % **elle est recommandée.** »⁹¹.

Le module de l'analyseur permettra d'identifier le « *bon* » groupe verbal, plus précisément *est recommandée*.

Configuration 7 : La phrase contient deux opérateurs déontiques ; la virgule est positionnée après le premier opérateur déontique.

Le pré-traitement permet de segmenter la phrase à partir de la virgule positionnée après l'opérateur déontique *pouvoir* à la voix passive.

⁹¹ Les anaphores mériteraient des traitements spécifiques permettant de les remplacer par leur sens réel et associer ainsi les actions spécifiques des verbes déontiques identifiés. Cette reconnaissance d'anaphores faciliterait le traitement automatique des phrases dans le but d'en dériver des règles de décision par exemple.

6.3. Stratégie de pré-traitement pour l'analyse de « textes réels »

« En cas de ganglion pédiculaire envahi, si la résecabilité est de classe I, la chirurgie avec curage **ne peut être contre-indiquée**, mais cette décision **doit néanmoins s'intégrer** dans une approche multidisciplinaire. »

La première proposition correspond donc à :

« En cas de ganglion pédiculaire envahi, si la résecabilité est de classe I, la chirurgie avec curage **ne peut être contre-indiquée,**»

Et la seconde à :

« mais cette décision **doit néanmoins s'intégrer** dans une approche multidisciplinaire. »

Configuration 8 : Utilisation du point-virgule

Un autre cas qui rejoint dans son principe la segmentation utilisant la virgule comme marqueur de segmentation, est l'utilisation du point virgule. Dans l'exemple ci-dessous, le point virgule correspond au marqueur de segmentation.

« L'association statine + fénofibrate **est classiquement déconseillée ; elle peut se discuter** après avis spécialisé. »

La première proposition est :

« L'association statine + fénofibrate **est classiquement déconseillée ;**»

Et la seconde proposition :

« elle **peut se discuter** après avis spécialisé. »

Configuration 9 : Utilisation des deux-points

Les deux-points peuvent introduire une liste énumérative, mais également une spécification de ce qui précède comme illustrée par l'exemple ci-dessous. Les deux-points correspondent donc, de la même manière que la virgule ou le point-virgule, à un marqueur de segmentation.

« Cette situation **ne doit être envisagée** qu'en cas d'intolérance avérée à la metformine (troubles digestifs) prescrite de façon adéquate ou de contre-indication (exceptionnelles dans un tel contexte) **⋮** dans cette rare situation et en présence d'un surpoids une des deux glitazones (Thiazolidine diones) **sera prescrite** en respectant leurs contre-indications (Accord professionnel). »

Les deux propositions issues de la segmentation sont donc :

- (1) « Cette situation **ne doit être envisagée** qu'en cas d'intolérance avérée à la metformine (troubles digestifs) prescrite de façon adéquate ou de contre-indication (exceptionnelles dans un tel contexte) **⚡** »
- (2) « dans cette rare situation et en présence d'un surpoids une des deux glitazones (Thiazolidine diones) **sera prescrite** en respectant leurs contre-indications (Accord professionnel). »

En revanche, les énumérations constituent un cas particulier pour le pré-traitement. Par exemple,

« Le diagnostic d'AVC, évoqué devant des troubles neurologiques focaux ou des troubles de la vigilance d'installation soudaine, **doit être confirmé** par l'imagerie cérébrale :

- L'imagerie cérébrale (IRM, scanner) **est nécessaire** sans délai pour affirmer le diagnostic d'AVC. Actuellement l'IRM-séquences écho-planar (méthode permettant d'obtenir une coupe IRM dans un temps d'acquisition de l'ordre de 100 ms) est l'examen le plus performant car elle permet le diagnostic très précoce
...
- Le scanner cérébral sans injection de produit de contraste permet le diagnostic en urgence d'hémorragie cérébrale. Les signes précoces d'ischémie cérébrale sont inconsistants et d'interprétation difficile.
- L'IRM avec séquences d'angioRM veineuse est l'examen non invasif le plus performant pour le diagnostic de thrombose veineuse cérébrale. »

Un module du pré-traitement décompte les occurrences des signes de ponctuation spécifiques aux énumérations (les deux-points suivis de tirets). Dans ce cas, le pré-traitement consiste à segmenter la phrase à partir du deux-points considéré comme marqueur de segmentation. Ce qui correspond à la sous-chaîne suivante :

« Le diagnostic d'AVC, évoqué devant des troubles neurologiques focaux ou des troubles de la vigilance d'installation soudaine, **doit être confirmé** par l'imagerie cérébrale :

Étant donné que cette sous-chaîne est une généralisation des spécifications énumérées⁹², nous considérons pour l'ensemble des traitements des énumérations que la seconde sous-chaîne ne sera pas soumise à l'analyseur, plus précisément :

- « L'imagerie cérébrale (IRM, scanner) **est nécessaire** sans délai pour affirmer le diagnostic d'AVC. Actuellement l'IRM-séquences écho-planar (méthode permettant d'obtenir une coupe IRM dans un temps d'acquisition de l'ordre de 100 ms) est l'examen le plus performant car elle permet le diagnostic très précoce
...

⁹² Nous prenons position quant au traitement des énumérations qui a été justifiée durant l'évaluation par les experts.

- Le scanner cérébral sans injection de produit de contraste permet le diagnostic en urgence d'hémorragie cérébrale. Les signes précoces d'ischémie cérébrale sont inconsistants et d'interprétation difficile.
- L'IRM avec séquences d'angioRM veineuse est l'examen non invasif le plus performant pour le diagnostic de thrombose veineuse cérébrale. »

6.3.5 Pré-traitements des phrases sans signes de ponctuation

Ce pré-traitement concerne les phrases contenant le signe de ponctuation final et deux opérateurs déontiques. Les signes de ponctuation ne peuvent donc pas être des marqueurs de segmentation pour ce type de phrase. Les marqueurs de segmentation peuvent alors être des marqueurs textuels, comme par exemple *mais* dans la phrase suivante.

« Des explorations complémentaires des gros vaisseaux, à commencer par l'exploration échodoppler, **ne sont pas recommandées** à titre systématique mais **seront proposées** en cas d'anomalies à l'examen clinique. »

Les deux propositions issues de la segmentation sont :

- (1) « Des explorations complémentaires des gros vaisseaux, à commencer par l'exploration échodoppler, **ne sont pas recommandées** à titre systématique mais »
- (2) « **seront proposées** en cas d'anomalies à l'examen clinique. »

D'autres marqueurs de segmentation sont recherchés comme des connecteurs, par exemple la conjonction de coordination *et*.

Dans une phrase ne contenant ni marqueur textuel ni signe de ponctuation mais contenant deux opérateurs déontiques, le marqueur de segmentation correspond au dernier terme de l'opérateur déontique. Par exemple,

« Lorsque cette classe thérapeutique **est mal tolérée ou contre-indiquée** un inhibiteur des alphasglucosidases **pourra être prescrit** (accord professionnel). »

Le verbe conjugué *contre-indiquée* correspond au marqueur de segmentation, les deux propositions sont donc :

- (2) « Lorsque cette classe thérapeutique **est mal tolérée ou contre-indiquée** »
- (3) « un inhibiteur des alphasglucosidases **pourra être prescrit** (accord professionnel). »

Chapitre 7

Fonctionnement de l'analyseur & intégration dans G-DEE

Ce chapitre décrit le fonctionnement de l'analyseur ainsi que l'algorithme dédié à l'extraction et à l'affichage du texte analysé, puis balisé. Ce chapitre fait le lien entre l'analyse linguistique et les aspects documentaires, représentés par le balisage des textes et différents traitements dérivés de ce balisage, qu'ils soient visuels comme les modalités d'affichage et de mise en valeur de segments textuels, ou conceptuels, comme la dérivation automatique de connaissances formalisées. Nous présentons donc les techniques utilisées basées essentiellement sur XML et les feuilles de style associées XSL (eXtensible Stylesheet Language) ainsi que les processeurs XSLT (XSL Transformations) permettant d'interpréter ces données. Les différents traitements présentés dans ce chapitre permettent de surligner le texte correspondant à des recommandations, d'identifier certaines informations du texte pour dériver de manière automatique des éléments GEM, ou des règles de décision.

7.1 Fonctionnement de l'analyseur

A la suite du pré-traitement décrit au chapitre précédent (en chaînes de caractères) chaque phrase d'un Guide de Bonnes Pratiques est analysée selon un processus (cf. Figure 7.81) que nous décrivons dans cette section. Nous utiliserons la phrase suivante pour illustrer l'ensemble des traitements :

« La radiothérapie n'est pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux). »

7.1.1 Étapes préliminaires

Les premières étapes de l'exécution du programme correspondent au chargement en mémoire des ressources linguistiques nécessaires à l'analyse. Tout d'abord, le contenu du fichier texte décrivant les motifs syntaxiques des opérateurs déontiques (Figure 7.77) nécessaires pour instancier les automates, est transféré dans une table de hachage⁹³.

```
{ [[pp_sing][aux_sing]]
[[op_deont_plur][verbe_aux][pp_assoc_plur]]
[[aux_plur prep verbe_inf]]
[[pronom neg verbe_inf_aux negp adj]]
[[neg_n][op_deont_plur][verbe_aux]]
[[neg_n aux_sing negp pp_sing]]
[[neg negp pp_sing adj]]
[[aux_sing][pp_sing prep verbe_inf]]
[[neg negp][pp_sing][prep verbe_inf]]
[[op_deont_sing][verbe_inf]]
[[neg_n op_deont_plur][verbe_aux]]
[[pronom op_deont_sing][verbe_aux][pp_assoc_sing]]
[[pronom op_deont_sing][verbe_aux][pp_sing]]
[[pronom op_deont_sing verbe_inf]]
[[aux_sing pp_sing][verbe_inf_aux pp_assoc_sing]]
[[pronom neg verbe_inf_aux negp adj prep verbe_inf]]
[[neg_n op_deont_plur][negp][verbe_inf]]
[[aux_sing pp_sing]]
[[op_deont_sing verbe_inf prep adj]]
[[op_deont_sing verbe aux verbe inf][prep verbe inf]] ... }
```

Figure 7.77 – Activation des motifs syntaxiques pour l'instanciation des automates.

Nous avons défini un formalisme de représentation de motifs syntaxiques permettant le stockage dans un fichier texte tout en étant lisible. Le double crochet ouvrant ([[) correspond à l'initialisation de l'automate et le double crochet fermant (]]) correspond à la fin de son instanciation. Le crochet fermant suivi d'un crochet ouvrant (] [) permet de définir une fenêtre d'analyse ; dans le cas des motifs syntaxiques décrivant les opérateurs déontiques, cette fenêtre est constituée de 4 mots.

⁹³ Une table de hachage est caractérisée par la paire {clé ; valeur} ce qui permet d'exécuter des algorithmes de recherche plus rapide qu'une recherche classique.

De la même manière, l'instanciation des catégories grammaticales (Figure 7.78) est identique à celle décrite précédemment pour les opérateurs déontiques. Néanmoins, cette grammaire permet d'instancier chaque motif d'opérateurs déontiques lors de l'exécution de l'analyseur.

```
{utilisés=pp_assoc_plur, recommander=verbe_inf, recommandée=pp_sing,
devraient=op_deont_plur, il=pronom, doivent=op_deont_plur,
conseillées=pp_plur, remplacer=verbe_inf, peuvent=op_deont_plur,
est=aux_sing, recommandées=pp_plur, pourra=op_deont_sing,
lorsque=cond, seront=aux_plur, lors=cond, utilisées=pp_assoc_plur,
,=virg, doit=op_deont_sing, de=prep, recommandés=pp_plur,
puissent=op_deont_plur, devra=op_deont_sing, conseillés=pp_plur,
pas=negp, ne=neg_n, devront=op_deont_plur, associer=verbe_inf,
pourront=op_deont_plur, pourrait=op_deont_sing,
devrait=op_deont_sing, être=verbe_aux, conseillé=pp_sing,
recommandé=pp_sing, pourraient=op_deont_plur,
utilisée=pp_assoc_sing, sera=aux_sing, utilisé=pp_assoc_sing,
peut=op_deont_sing, en cas de=cond, recommande=op_deont_sing,
recommandent=op_deont_plur, sont=aux_plur, conseillée=pp_sing ... }
```

Figure 7.78 – Instanciation des catégories grammaticales dans une table de hachage.

Dans le but de reconnaître les formes infinitives englobant l'opérateur déontique, des motifs syntaxiques sont définis et permettent l'analyse de la phrase pour la recherche et le balisage des formes infinitives dans la phrase balisée. Ces motifs sont également décrits dans un fichier texte, et transférés dans une table de hachage à l'exécution du programme (Figure 7.79).

```
[[<FrontScope>][</FrontScope>][<OpReco>][verbe_inf][</OpReco>][<BackScope>][</BackScope>]],
[[<FrontScope>][</FrontScope>][<OpReco_P>][verbe_inf][</OpReco_P>][<BackScope>][</BackScope>]]
```

Figure 7.79 – Instanciation des formes infinitives

Concernant la reconnaissance de formes conditionnelles, les différents motifs syntaxiques de ces formes sont également décrits dans un fichier texte (Figure 7.80). Ces motifs reprennent le balisage de la phrase en termes de *frontscope* et de *backscope* ainsi que les balises de l'opérateur déontique, *<Op_Reco>*, *<Op_Reco_P>*, *</Op_Reco>*, et *</Op_Reco_P>*.

```
{ [ <FrontScope>][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][cond][virg][cond][</BackScope>]]
[[<FrontScope>][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][</BackScope>][<FrontScope>][cond]
[</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][</BackScope>]]
[[<FrontScope>][cond][virg][</FrontScope>][<OpReco>][</OpReco>][<BackScope>][</BackScope>]]
[[<FrontScope>][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][cond][virg][</BackScope>]]
[[<FrontScope>][cond][virg][</FrontScope>][<OpReco>][</OpReco>][<BackScope>][cond][virg][cond]
[</BackScope>]]
[[<FrontScope>][cond][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][</BackScope>][<FrontScope>]
[</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][</BackScope>]]
[[<FrontScope>][</FrontScope>][<OpReco>][</OpReco>][<BackScope>][cond][</BackScope>]]
[<FrontScope>][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][cond][</BackScope>][<FrontScope>]
[cond][</FrontScope>][<OpReco_P>][</OpReco_P>][<BackScope>][</BackScope>]]
... )
```

Figure 7.80 – Instanciation des formes conditionnelles.

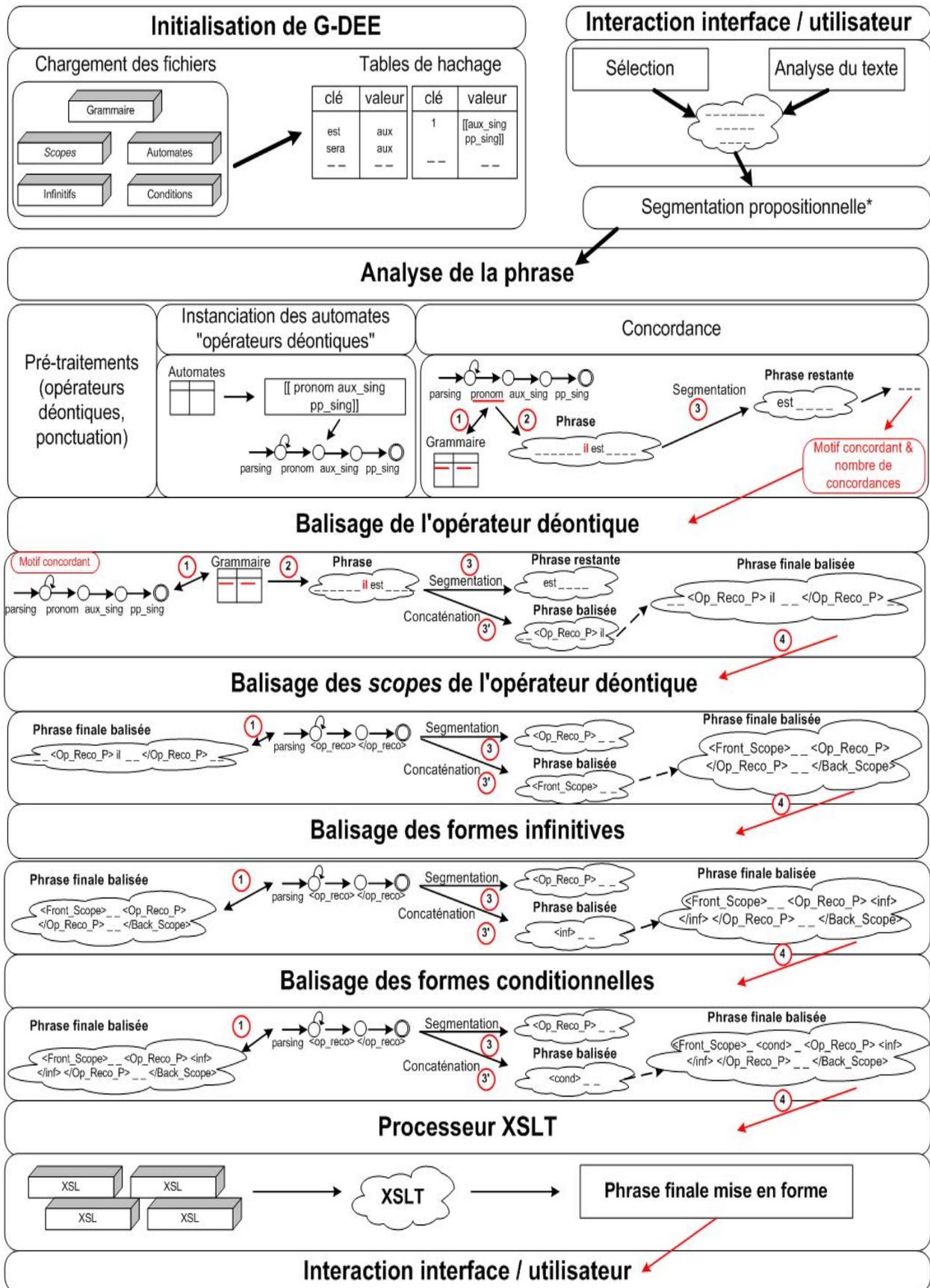


Figure 7.81 – Représentation générale du processus de balisage d'une phrase dans G-DEE.

7.1.2 Balisage des opérateurs déontiques

Chaque phrase du Guide de Bonnes Pratiques est analysée en séquence par sélection de l'automate le plus spécifique⁹⁴.

Pour chaque phrase à analyser, le module de reconnaissance des opérateurs déontiques est exécuté et déclenche les opérations suivantes :

- (i) activation de l'automate par un motif de la table de hachage
- (ii) analyse de la phrase par instanciation du motif

Chaque motif syntaxique décrivant les opérateurs déontiques et contenu dans la table de hachage est testé, et le nombre d'éléments concordant du motif avec la phrase est calculé dans le but de gérer les problèmes de conflits entre plusieurs motifs syntaxiques (instanciation possible de plusieurs motifs), et plus particulièrement pour faire correspondre le motif le plus pertinent à la phrase.

L'exemple ci-dessus `[[aux_plur][pp-plur adv verbe_inf]]` illustre le processus d'instanciation d'un motif syntaxique sur la phrase à analyser. Après activation de l'automate, le premier élément du motif `aux_plur` qui correspond à la catégorie des auxiliaires au pluriel, comme « sont » et « seront », est instancié. La phrase analysée est ensuite parcourue à la recherche de ces auxiliaires. Dans l'exemple ci-dessous, aucun mot de la phrase ne correspond aux valeurs de `aux_plur`. Le motif considéré `[[aux_plur][pp-plur adv verbe_inf]]` ne peut donc pas être concordant avec la phrase analysée. Un autre motif de la table de hachage est analysé.

Phrase à analyser :

La radiothérapie n'est pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

Motif syntaxique considéré : `[[aux_plur][pp-plur adv verbe_inf]]`

Activation du motif : `[[`

Instanciation de l'élément 'aux_plur' par :

- sont
- seront

Motif syntaxique non concordant [0 élément du motif retrouvé]

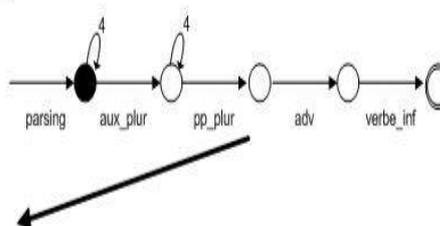


Figure 7.82 – Illustration de l'analyse de la phrase par instanciation de l'automate avec le motif `[[aux_plur][pp-plur adv verbe_inf]]`. Le motif n'est pas concordant avec la phrase.

⁹⁴ On entend ici par « pertinent » un motif qui est le plus concordant possible avec la phrase analysée (rappelons que le système ne procède pas à une analyse complète et que plusieurs motifs peuvent donc être instanciables sur une même phrase). L'identification d'un motif comprenant un auxiliaire et un participe passé est donc moins spécifique que celle d'un groupe verbal composé d'un auxiliaire, d'un participe passé suivi d'un adverbe et d'un verbe à l'infinitif.

L'exemple ci-dessous illustre l'instanciation de l'automate par le motif syntaxique « [[neg][negp pp_sing]] ».

De manière similaire au traitement précédent, l'analyseur va d'abord chercher une correspondance avec la valeur de *neg* dans la grammaire (correspondant à la négation de l'auxiliaire *être*, i.e « n'est »). La recherche de cet élément aboutit (Figure 7.83), on poursuit donc l'analyse sur le reste de la phrase : « *pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, jambes, organes génitaux).* »

Instanciation de l'élément 'neg' par :

- n'est

Phrase restante à analyser : pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

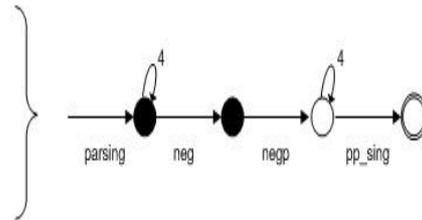


Figure 7.83 – Illustration de l'analyse représentant l'instanciation du premier élément du motif par *n'est*.

L'élément suivant (négation *pas*) est aussitôt reconnu dans la phrase, et l'analyse se poursuit à partir de cet élément, sur le reste de la phrase (Figure 7.84).

Instanciation de l'élément 'negp' par :

- pas

Recherche de l'élément avec une fenêtre de recherche de 0.

Phrase restante à analyser : pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

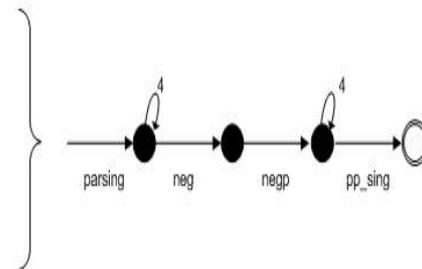


Figure 7.84 – Illustration de l'analyse représentant l'instanciation du motif par *pas*.

L'élément suivant du motif, «]] », correspond à une fenêtre d'analyse de 4 mots pour la recherche de l'élément suivant (*pp_sing* un participe passé à la 3^{ème} personne du singulier) (Figure 7.85). Cette fenêtre d'analyse permet de contourner différentes séquences, comme les locutions adverbiales⁹⁵. La catégorie du participe passé est donc ensuite recherchée sur le reste de la phrase, pour chacun des verbes du vocabulaire (qui sont bien sûr des verbes déontiques, comme *conseiller* ou *recommander*, à ce stade des motifs analysés). C'est le verbe *recommander* qui est reconnu, ce qui conclut l'instanciation du motif.

⁹⁵ On aura remarqué que la catégorie adverbiale est pourtant explicitement représentée dans certains motifs. Nous considérons que, si la plupart des locution adverbiales apportent peu d'information ("non plus", "de plus en plus", "pas vraiment"), la reconnaissance de certains adverbes (comme "toujours" et "jamais") peut contribuer à l'interprétation des recommandations.

Instanciation de l'élément 'pp_sing' par :

- conseillé
- déconseillé
- recommandé
- recommandée

Recherche de l'élément avec une fenêtre de recherche de 4.

Phrase restante à analyser : chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

Motif syntaxique concordant [3 éléments du motif retrouvés]

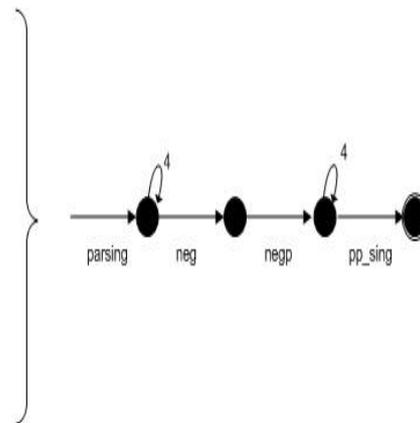
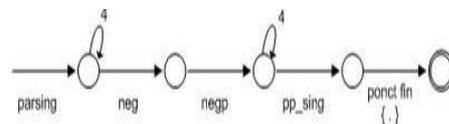


Figure 7.85 – Illustration de l'analyse représentant l'instanciation de l'élément du motif *pp_sing* par les verbes déontiques de la grammaire, et indiquant le nombre d'éléments concordant avec la phrase.

Le motif syntaxique ayant été reconnu, le processus de balisage de l'opérateur déontique correspondant peut être réalisé, comme le décrit la figure ci-dessous. La balise de début de l'opérateur déontique précède le premier élément du motif, i.e. *neg* (*n'est*), et la balise de fin succède au dernier élément du motif, i.e. *pp_sing* (*recommandée*).

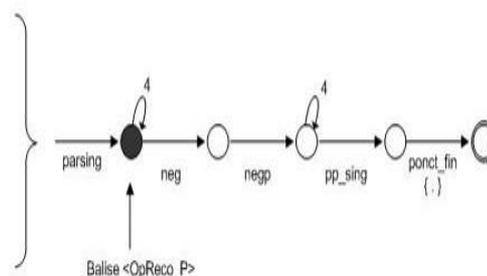
Motif syntaxique concordant : `[[neg negp][pp_sing]]`



Instanciation du premier élément du motif 'neg' par n'est

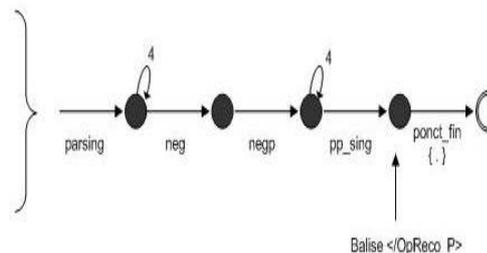
Balise de début de l'opérateur déontique : La radiothérapie `<Op_Reco_P>` n'est

Phrase restante à baliser : pas non plus recommandée chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).



Activation du dernier élément du motif 'pp_sing' par recommandée

Balise de fin de l'opérateur déontique : La radiothérapie `<Op_Reco_P>` n'est pas non plus recommandée `</OpReco_P>`



Phrase balisée : La radiothérapie `<Op_Reco_P>` n'est pas non plus recommandée `</OpReco_P>` chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiformes, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

Figure 7.86 – Illustration de l'analyse correspondant au balisage de l'opérateur déontique en fonction du motif reconnu précédemment.

7.1.3 Balisage des *scopes*

Le processus de balisage des *scopes* de l'opérateur déontique est présenté par la Figure 7.87. L'étape préalable nécessaire à la réalisation de ce processus est le balisage de l'opérateur déontique. En effet, le balisage des *scopes* est fonction des balises de l'opérateur déontique.

Motif syntaxique concordant : `[[neg neg][pp_sing]]`

Activation du premier élément de la phrase balisée :
`<Front_Scope>`

Activation du premier élément du motif 'neg' par n'est

1^{ère} partie de la phrase à baliser : La radiothérapie
`<Op_Reco_P>` n'est

2^{ème} partie de la phrase : pas non plus recommandée
`</Op_Reco_P>` chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

Segment précédant la balise de début d'opérateur déontique :
 La radiothérapie

Balise de fin pour le FrontScope : La radiothérapie
`</FrontScope>`

Fin de balisage de la phrase : `<Front_Scope>` La radiothérapie
`</FrontScope>` `<OpReco_P>` n'est

2^{ème} partie de la phrase à baliser : pas non plus recommandée
`</Op_Reco_P>` chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux).

Ajout de la balise de début de Backscope : pas non plus recommandée
`</OpReco_P>` `<Back_Scope>`

Ajout de la balise de fin de Backscope : recommandée
`</OpReco_P>` `<Back_Scope>` chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux)
`</Back_Scope>`.

Phrase balisée :

`<Front_Scope>` La radiothérapie `</Front_Scope>` `<OpReco_P>` n'est pas non plus recommandée `</OpReco_P>`

`<Back_Scope>` chez les sujets de moins de 60 ans, comme traitement des CBC sclérodermiques, sur certaines zones (oreilles, mains, pieds, jambes, organes génitaux) `</Back_Scope>`.

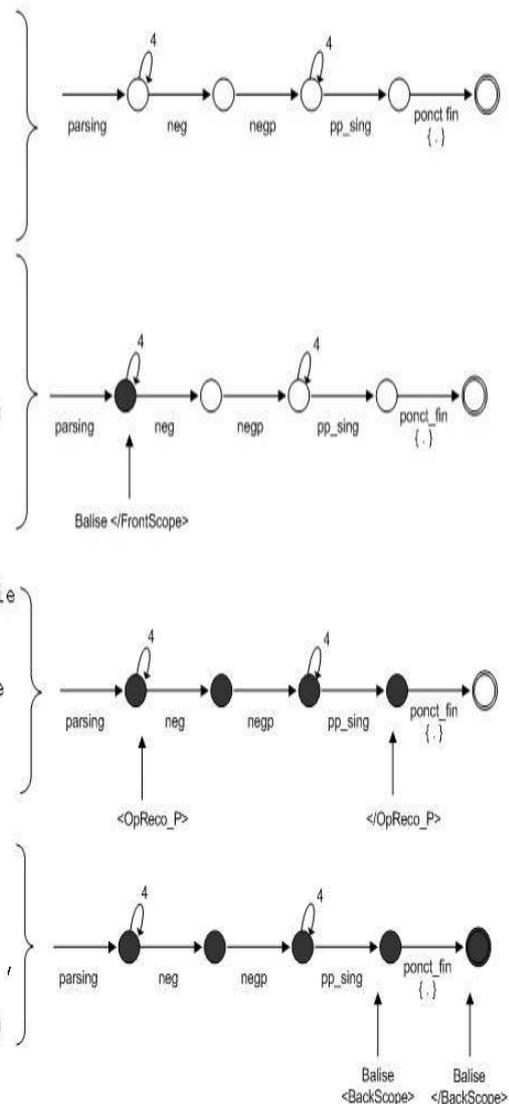


Figure 7.87 – Balisage du *front-scope* et du *backscope* en fonction du motif reconnu.

7.1.4 Balisage des formes infinitives

L'exemple précédent ne contient pas de formes infinitives dans le groupe verbal déontique et ne permet donc pas de décrire le fonctionnement de l'analyseur sur les formes infinitives des verbes déontiques. L'exemple choisi pour représenter cette étape est :

Si le diabète est diagnostiqué chez un patient âgé, un objectif de HbA1c comprise entre 6,5% et 8,5% peut servir de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel).

La phrase est analysée selon le processus décrit précédemment pour reconnaître et baliser l'opérateur déontique, le *front-scope* et le *back-scope*.

La phrase considérée est composée du groupe verbal déontique *peut servir* et le balisage est généré en fonction du motif reconnu « [[op_deont_sing verbe_inf]] », qui correspond à l'opérateur déontique à la 3^{ème} personne du singulier suivi d'un verbe à l'infinitif.

La phrase balisée selon le motif reconnu est :

<FrontScope> Si le diabète est diagnostiqué chez un patient âgé, un objectif de HbA1c comprise entre 6,5% et 8,5% </FrontScope> <OpReco> peut servir </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.

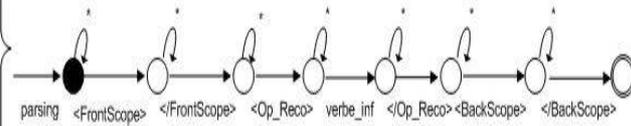
La reconnaissance de la forme infinitive dans le groupe verbal déontique est assez triviale puisque l'analyseur va rechercher uniquement dans le segment textuel de l'opérateur déontique une forme infinitive d'un verbe présent dans la grammaire. En général, ce verbe à l'infinitif indique une action à réaliser, par exemple « *peut servir* ». La figure ci-dessous décrit le processus de recherche du segment à considérer. L'analyseur instancie ensuite l'automate par le verbe à l'infinitif.

Motif syntaxique considéré :

[[<FrontScope>][</FrontScope>][<OpReco>][verbe_inf][</OpReco>][<BackScope>][</BackScope>]]

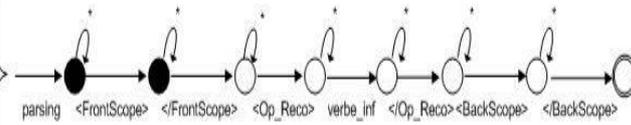
Activation de l'élément : <FrontScope>

Phrase restante à analyser : Si le diabète est diagnostiqué chez un patient âgé , un objectif de HbA1c comprise entre 6,5% et 8,5%
</FrontScope> <OpReco> peut servir </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



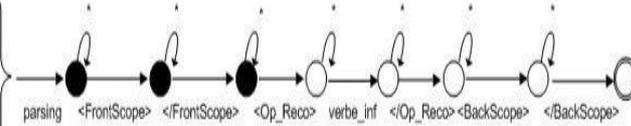
Activation de l'élément : </FrontScope>

Phrase restante à analyser : <OpReco> peut servir </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



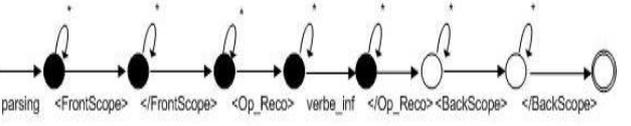
Activation de l'élément : <OpReco>

Phrase restante à analyser : peut servir </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



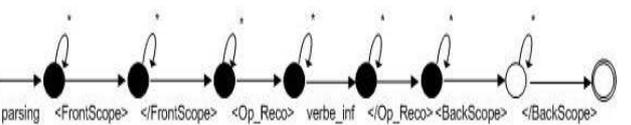
Activation de l'élément `verbe_inf` par servir

Phrase restante à analyser : servir </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



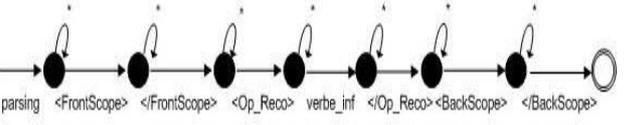
Activation de l'élément `</OpReco>`

Phrase restante à analyser : <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



Activation de l'élément `<BackScope>`

Phrase restante à analyser : de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.



Instanciation de l'élément `</BackScope>`

Motif syntaxique concordant
[7 éléments du motif retrouvés]

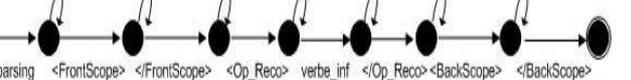


Figure 7.88 – Reconnaissance de la forme infinitive englobée dans la balise de l'opérateur déontique.

7.1. L'analyseur

La figure ci-dessous représente le processus de balisage de la forme infinitive englobée dans le groupe verbal déontique.

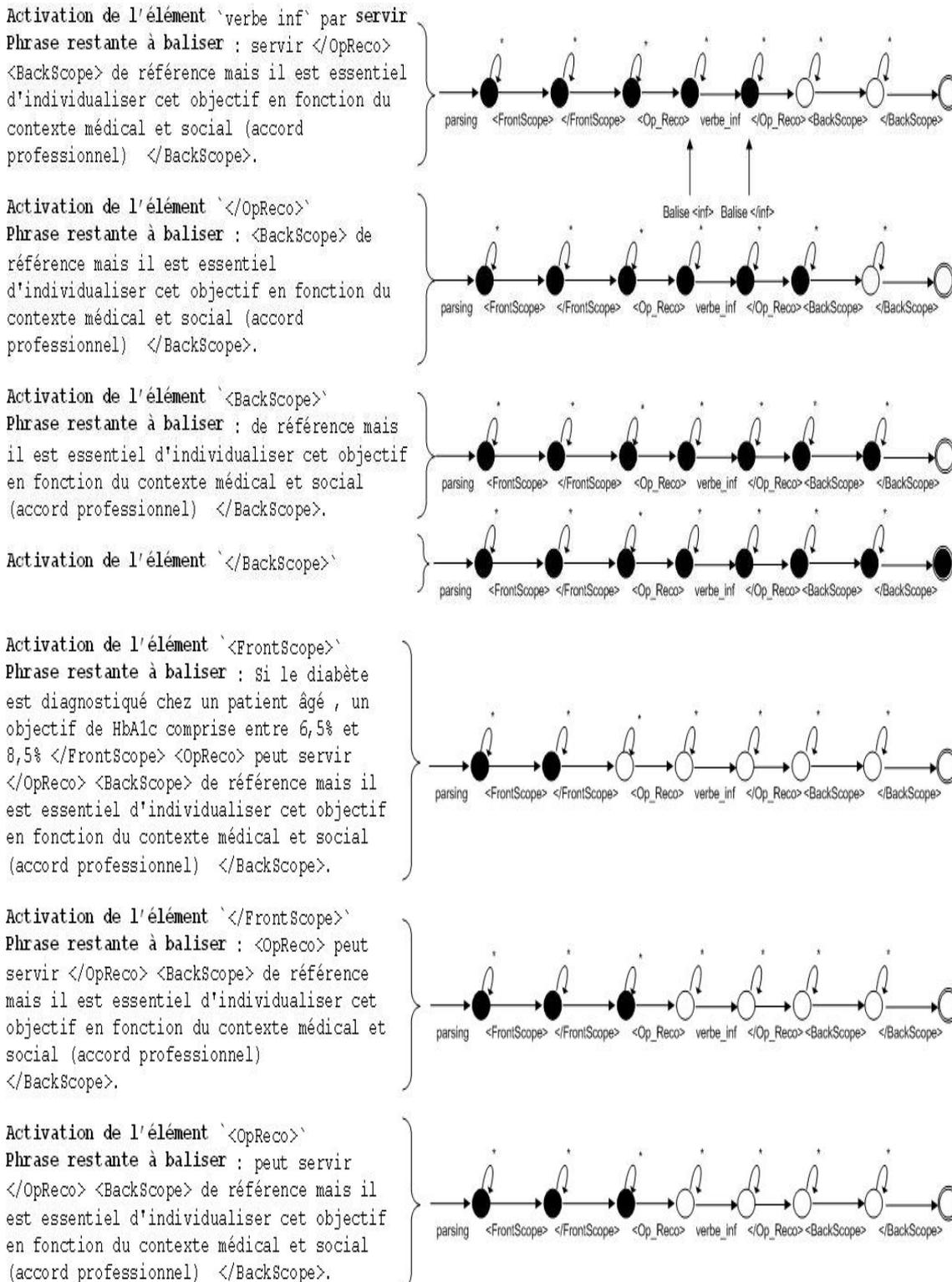


Figure 7.89 – Balisage de la forme infinitive englobée dans le groupe verbal déontique.

7.1.5 Balisage des connecteurs de condition

La première étape consiste à reconnaître si le *front-scope* ou le *back-scope* (ou les deux) contiennent des connecteurs de condition (Figure 7.90).

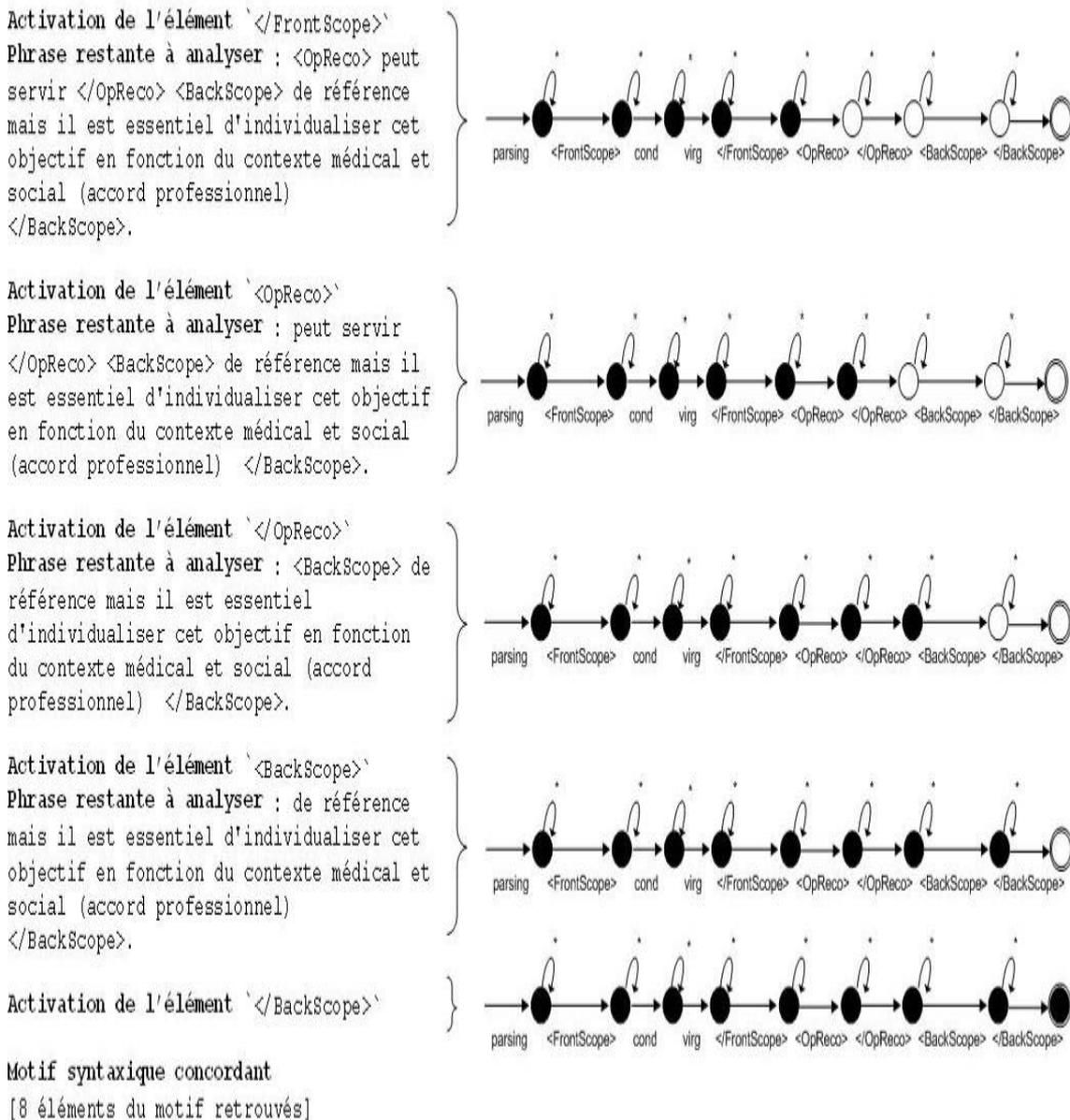


Figure 7.90 – Reconnaissance de connecteurs de condition dans les *scopes* de l'opérateur déontique.

La seconde étape consiste à baliser les connecteurs de condition (Figure 7.91).

Phrase à baliser :

<FrontScope> Si le diabète est diagnostiqué chez un patient âgé , un objectif de HbA1c comprise entre 6,5% et 8,5% </FrontScope> <OpReco> peut <inf> servir </inf> </OpReco> <BackScope> de référence mais il est essentiel d'individualiser cet objectif en fonction du contexte médical et social (accord professionnel) </BackScope>.

Motif syntaxique considéré pour le balisage :

[[<FrontScope>] [<cond>] [virg] [</FrontScope>] [<OpReco>] [</OpReco>] [<BackScope>] [</BackScope>]]

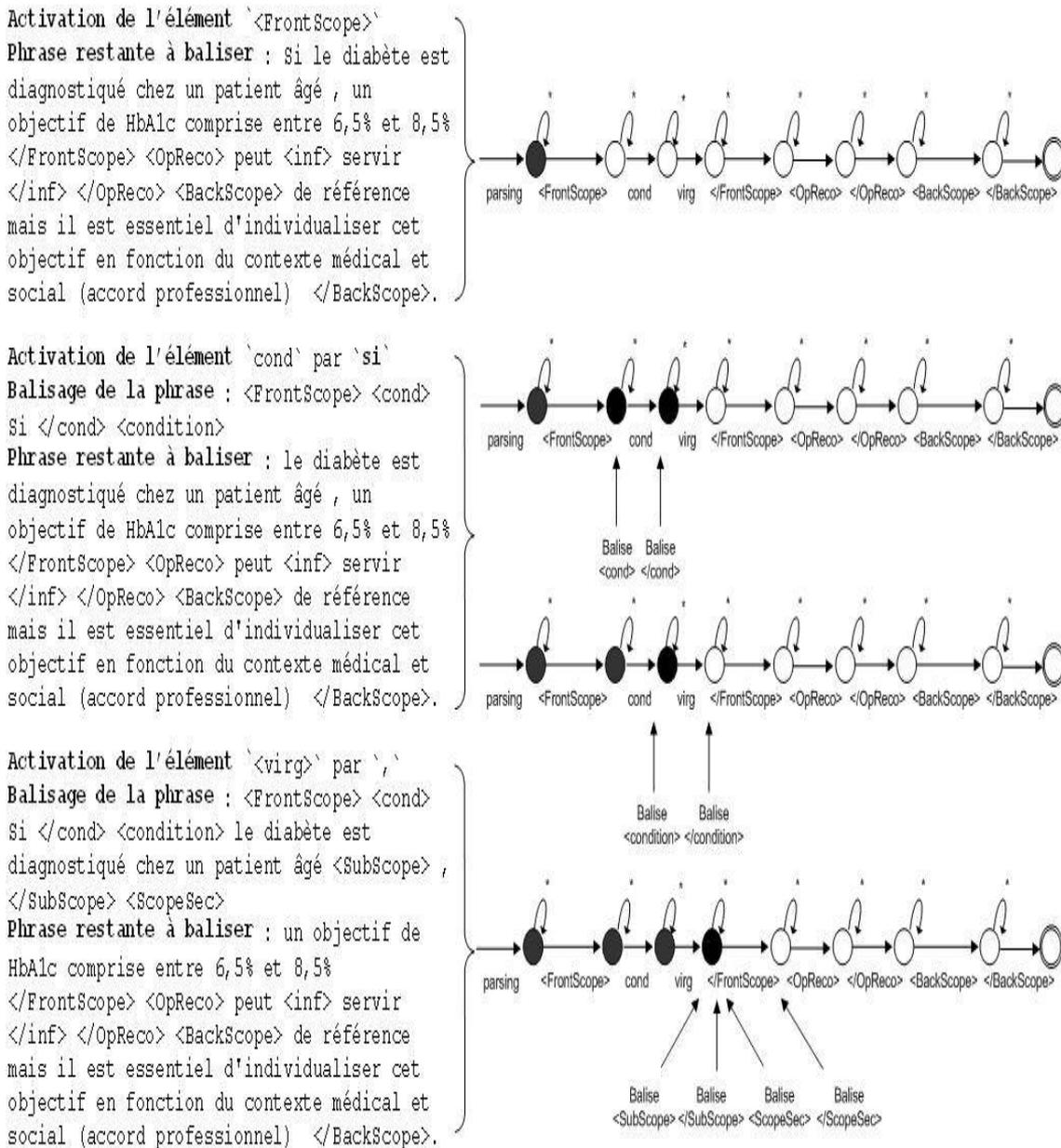


Figure 7.91 – Balisage des connecteurs de condition et du segment correspondant à la condition.

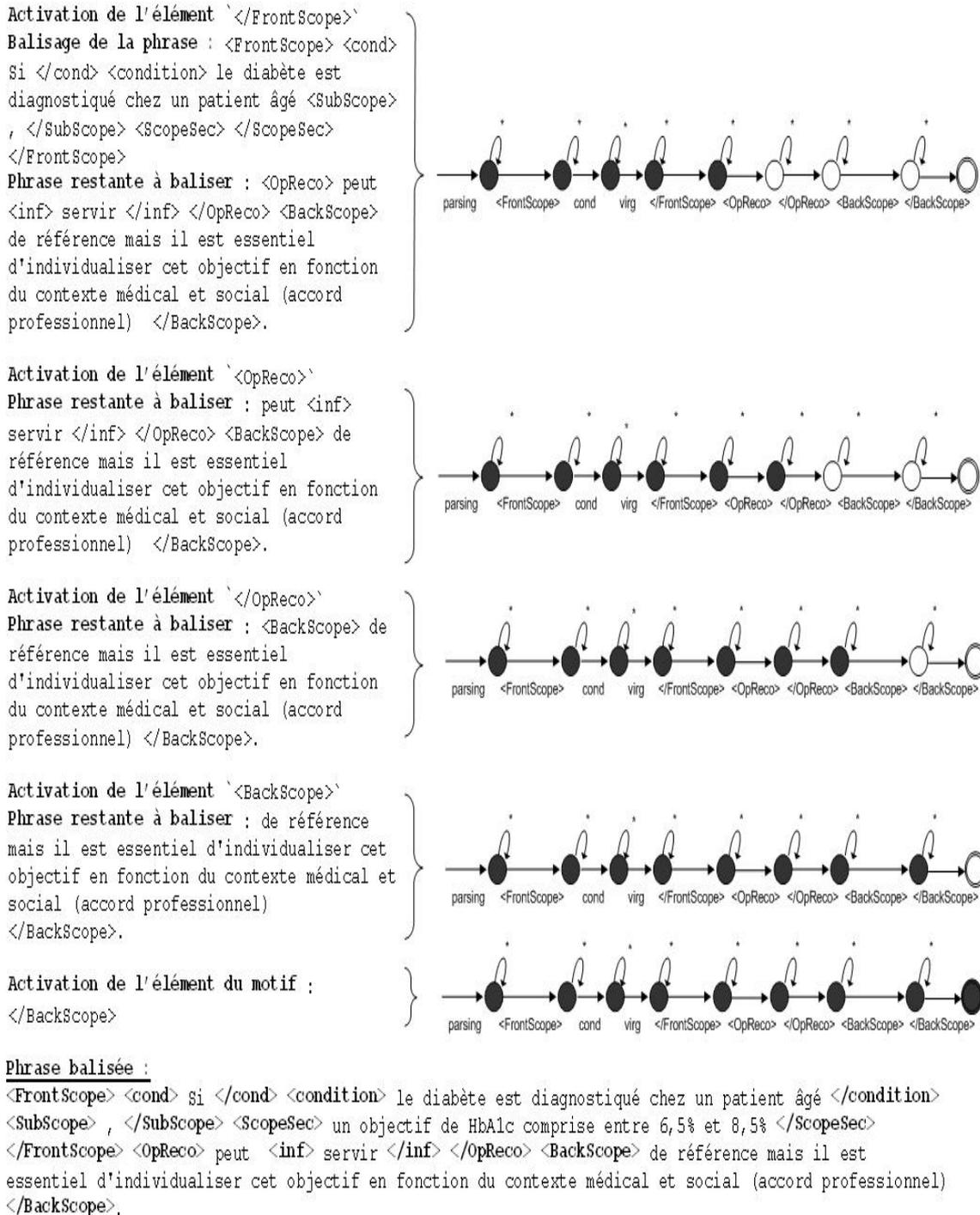


Figure 7.92 – Suite de la Figure 7.91 permettant de baliser les connecteurs de condition et les segments textuels correspondant à la condition de la phrase.

La structure de la phrase balisée en présence d'un marqueur de condition dans le *back-scope*, par exemple en cas de, correspond donc à :

```
<FrontScope> </FrontScope> <OpReco_P> </OpReco_P> <BackScope> <cond>
</cond> <condition> </condition> <SubScope> </SubScope> <ScopeSec>
</ScopeSec> </BackScope>.
```

Nous avons introduit de nouvelles balises, *<SubScope>* et *<ScopeSec>* pour caractériser chacun des segments de la phrase, et en particulier, identifier à l'intérieur du *front-scope* les différents éléments de la condition. Ce balisage additionnel est utilisé en particulier pour la génération de règles de décision à partir de texte balisé, comme décrit plus loin.

7.2 Performance et Complexité de l'algorithme

L'efficacité des algorithmes de traitement est un facteur limitant de l'analyse de documents, en particulier dans la perspective d'une intégration dans un système interactif. La figure ci-dessous représente la complexité de l'algorithme élaboré pour l'analyse du Guide de Bonnes Pratiques de la « prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral » (ANAES, 2002) contenant 490 lignes (~5 273 mots).

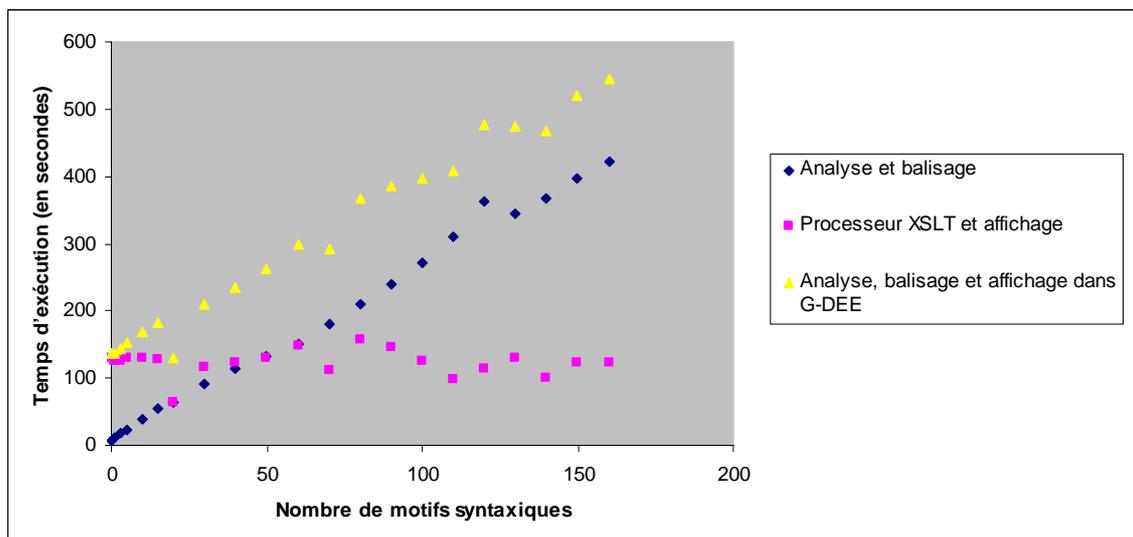


Figure 7.93 – Représentation graphique de la complexité de l'algorithme implémenté.

Ce graphique étudie le temps d'exécution du programme (en secondes) en fonction du nombre de motifs syntaxiques utilisés pour la reconnaissance des opérateurs déontiques (en abscisse). Il suggère une complexité linéaire en fonction de la « grammaire » des opérateurs déontiques. La courbe jaune représente le temps d'exécution de G-DEE, plus précisément l'analyse du texte entier, le balisage des phrases reconnues comme des recommandations et l'affichage consistant à surligner les opérateurs déontiques et ses *scopes*. La courbe rose

correspond au temps d'exécution du processeur XSLT et de l'affichage du texte balisé ; ce temps est plus au moins constant, car il ne dépend naturellement pas du nombre total de motifs reconnaissables mais du nombre de motifs effectivement instanciés dans un document donné. La courbe bleue concerne le temps d'exécution de l'analyseur de G-DEE (reconnaissance et balisage des phrases reconnues comme des recommandations).

En conclusion, nous pouvons revenir sur nos hypothèses initiales quant aux techniques de traitement automatique du langage naturel à utiliser. Le problème de la reconnaissance d'expressions déontiques dans des textes libres se prête bien à l'analyse par des méthodes empiriques telles que celles développées en Extraction d'Informations. En définissant une base de motifs, nous disposons d'une certaine souplesse face à la variabilité des occurrences des expressions déontiques. Le prix à payer pour cette souplesse correspond au recouvrement entre motifs, qui va quelque peu allonger le temps de traitement. La performance obtenue reste néanmoins compatible avec nos objectifs initiaux, surtout en considérant le potentiel d'optimisation de certains aspects des algorithmes utilisés.

7.3 Intégration des fonctionnalités

La reconnaissance automatique d'expressions déontiques constitue la première étape d'une série de fonctions de structuration et d'affichage des Guides de Bonnes Pratiques. Ces fonctions se situent entièrement dans l'environnement XML au sens large, tant pour les techniques d'affichage que pour les approches plus spécifiques au domaine (comme GEM). Cette section est dédiée au processus d'interprétation des phrases balisées, plus spécifiquement au processeur XSLT⁹⁶ développé permettant de transformer des documents au format XML en HTML. Le but est de pouvoir personnaliser l'affichage du texte balisé par le surlignement en couleur par exemple, ou de dériver automatiquement les éléments nécessaires pour représenter le texte selon le modèle GEM ou sous la forme de règles de décision (Figure 7.94).

Dans le but de spécifier les modalités d'affichages, nous considérons d'abord qu'une recommandation peut être représentée sous la forme d'une règle constituée d'une partie condition et d'une partie action, suivant en cela (Shiffman et al., 2000). Dans cette optique, des règles basées sur la reconnaissance du *front-scope* et du *back-scope* peuvent être définies pour caractériser de manière automatique les éléments correspondant aux critères de décision et d'action d'une recommandation. Ce sont des règles de ce type que nous avons définies et intégrées dans des feuilles de style pour exécution par un processeur XSLT. Celui-ci intervient également lors de l'affichage du texte contenant des phrases surlignées en couleur, et permet de transformer le fichier XML structuré (le document balisé) en un fichier HTML pouvant être intégré dans G-DEE pour visualisation dans l'interface graphique.

⁹⁶ <http://www.w3.org/TR/xslt> (dernier accès le 10 juin 2006).

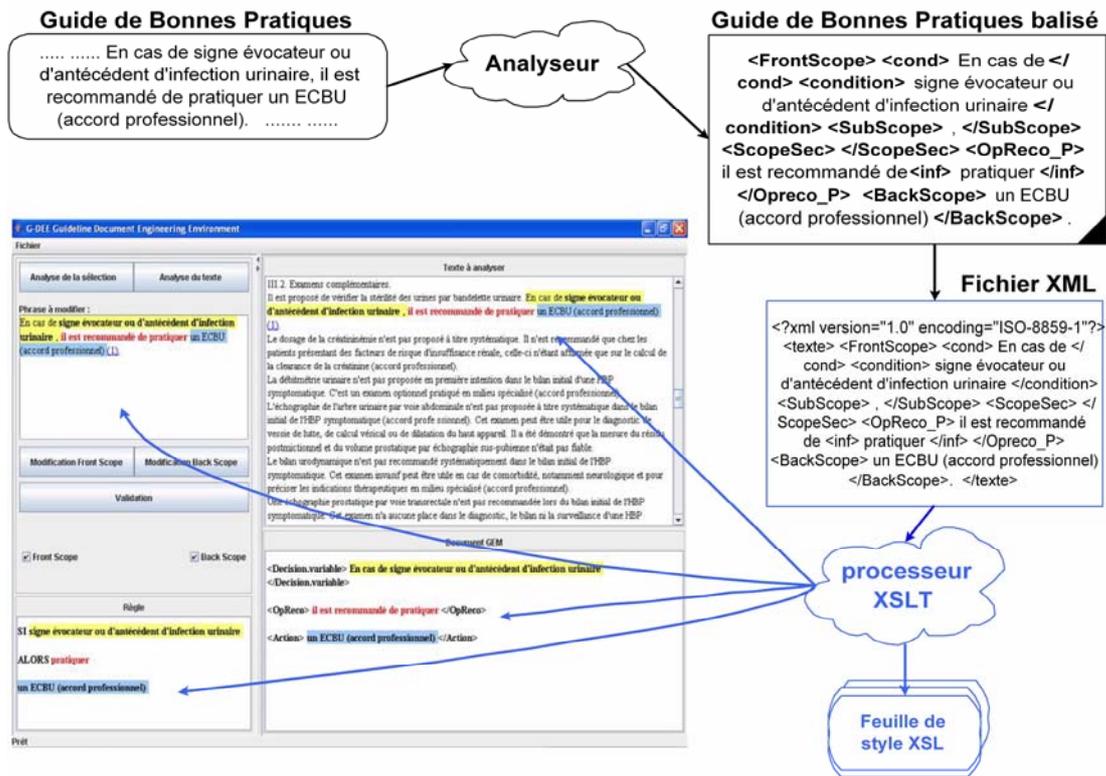


Figure 7.94 – Représentation du rôle du processeur XSLT, ses dépendances et l'exécution du processeur intervenant principalement pour l'affichage tel que le surlignement dans le texte.

7.3.1 Règles de transformation

Les segments textuels entourant l'opérateur déontique, plus particulièrement le texte correspondant au contenu du *front-scope* et du *back-scope* indiquent des conditions et des actions de l'application d'une recommandation.

Chaque recommandation détectée par G-DEE peut être représentée selon les éléments du modèle GEM (Shiffman et al., 2000), et sous la forme de règles de décision (Georg & Jaulent, 2005). Nous décrivons chaque configuration de recommandation par un ensemble d'exemples extraits de Guides de Bonnes Pratiques. La première configuration étudiée concerne les phrases à la voix passive représentée par les trois exemples ci-dessous (l'opérateur déontique est en rouge dans le texte).

*Exemple 1 : En cas de prescription d'un traitement par IEC ou ARAI, **il est recommandé de prescrire** un bilan associant kaliémie et créatininémie dans un délai de 7 à 15 j, après le début du traitement.*

*Exemple 2 : La recherche d'une hypotension orthostatique (chute de la PAS de plus de 20 mmHg et/ou de la PAD de plus de 10 mmHg, lors du passage en position debout), **est conseillée chez** tout hypertendu, en particulier **chez** le sujet de plus de 65 ans et le patient diabétique.*

*Exemple 3 : Une formation clinique en psycho-oncologie **doit être proposée et conseillée** aux cliniciens prenant en charge le suivi de femmes traitées pour un cancer du sein.*

En considérant le premier exemple, les segments du *front-scope* correspondent aux conditions, alors que dans le second exemple ils concernent le *back-scope*. La règle pour dériver de manière automatique des éléments GEM ou des règles de décision peut se définir ainsi :

- Lorsqu'un marqueur textuel indiquant une condition, par exemple en cas de, est un composant du *front-scope*, le *front-scope* correspond à la partie condition, et le *back-scope* à l'action.
- Lorsqu'un marqueur textuel indiquant une condition est un composant du *back-scope*, le *back-scope* correspond à la condition, et le *front-scope* à l'action.
- Si aucun marqueur textuel n'indique une condition : le *front-scope* correspond à l'action, et le *back-scope* à la condition.

La seconde configuration concerne les phrases à la voix active, comme celles représentées ci-dessous :

*Exemple 1 : **En cas de** maladie rénale ou d'insuffisance rénale, le contrôle de la pression artérielle **nécessite** le plus souvent une association thérapeutique.*

*Exemple 2 : Ce traitement par l'aspirine **ne doit être initié que lorsque** la PA est contrôlée car le risque d'hémorragie cérébrale est augmenté chez le patient hypertendu non contrôlé (Grade C).*

*Exemple 3 : Le traitement pharmacologique du patient diabétique de type 2 **devra souvent associer** de nombreuses médications pour obtenir les valeurs cibles recommandées pour chacun des facteurs de risque.*

Dans le premier exemple, les segments correspondant aux conditions concernent le *front-scope*, alors que dans le second elles impliquent le *back-scope*. La règle découlant de ces exemples est la suivante :

- Lorsqu'un marqueur textuel indiquant une condition est un composant du *front-scope*, le *front-scope* correspond à la condition, et le *back-scope* à l'action.
- Lorsqu'un marqueur textuel indiquant une condition se situe dans le *back-scope*, le *back-scope* correspond à la condition, et le *front-scope* à l'action.
- S'il n'existe aucun marqueur textuel indiquant une condition, alors le *front-scope* correspond à la condition, et le *back-scope* à l'action.

En résumé, la présence d'un marqueur textuel indiquant une condition dans un *scope* quelle que soit la voix de l'opérateur déontique implique que ce *scope* contienne la condition, et

inversement le *scope* complémentaire (au sens *front/back*) contient l'action. Concernant les phrases ne contenant pas de tels marqueurs, la voix de l'opérateur doit alors être prise en compte, et la règle est définie comme suit : « le *front-scope* d'une phrase active correspond à la condition et le *back-scope* à l'action, et inversement pour la voix passive ».

7.3.2 Structure du fichier XML du texte balisé

G-DEE permet le traitement direct des fichiers texte, néanmoins dans le souci d'afficher certains éléments surlignés, un module est dédié à la transformation du fichier texte en HTML (Figure 7.95). Le fichier source est également transformé en un fichier XML nécessaire au processeur XSLT pour interpréter le texte balisé par l'analyseur.

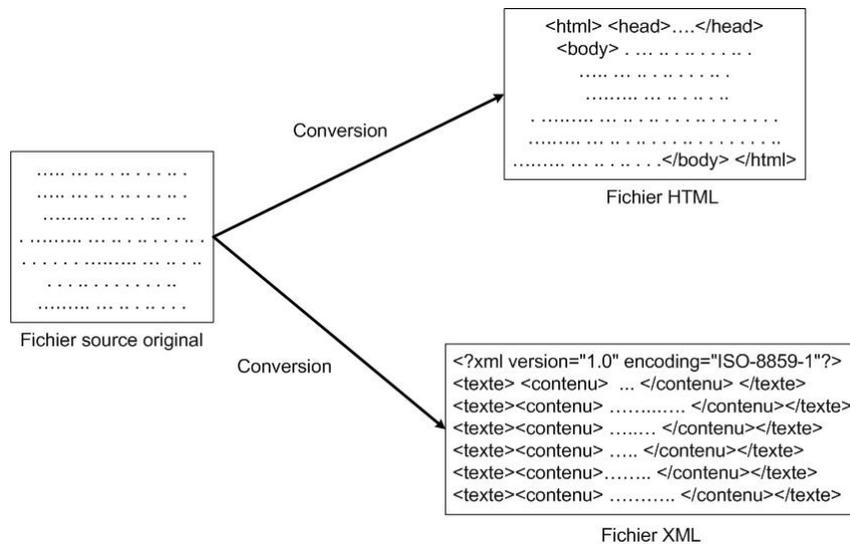


Figure 7.95 – Conversion du fichier texte source en fichiers HTML et XML nécessaires lors l'affichage dans G-DEE.

Un module est dédié au balisage de chaque phrase du texte source caractérisé par les balises *<texte>*, et *<contenu>*. Ces balises sont rendues nécessaires pour le traitement ultérieur du fichier XML par le processeur XSLT.

7.3.3 Interprétation de surface du texte balisé

7.3.3.1 Le processeur XSLT

Le processeur XSLT permet d'interpréter le document XML en fonction de feuilles de style définies pour générer un fichier HTML spécifique. Pour chaque fenêtre d'affichage de l'environnement, une feuille de style différente est activée. Le code du programme permettant de faire une telle transformation est représenté par la Figure 7.96. Les premières déclarations permettent de définir les fichiers concernés, tels que le fichier HTML généré en sortie d'exécution du programme, la feuille de style concernée et le document XML contenant le texte balisé. Le second bloc d'instructions permet d'instancier un parseur

SAX⁹⁷ pour analyser et interpréter les balises en fonction de la feuille de style, puis de transformer le fichier XML en un fichier HTML.

```

FileOutputStream out = new FileOutputStream("F:docHtmlSelect.html");
FileInputStream xslInput = new FileInputStream("F:style.xsl");
FileInputStream xmlInput = new FileInputStream("F:docxml.xml");

TransformerFactory tFactory = TransformerFactory.newInstance();
SAXTransformerFactory saxTFactory = ((SAXTransformerFactory)tFactory);
TemplatesHandler templatesHandler = saxTFactory.newTemplatesHandler();
XMLReader reader = XMLReaderFactory.createXMLReader();
reader.setContentHandler(templatesHandler);
reader.parse(new InputSource(xslInput));
Templates templates = templatesHandler.getTemplates();
TransformerHandler handler = saxTFactory.newTransformerHandler(templates);
reader.setContentHandler(handler);
reader.setProperty("http://xml.org/sax/properties/lexical-handler", handler);
Serializer serializer = SerializerFactory.getSerializer(OutputPropertiesFactory.getDefaultMethodProperties("html"));
serializer.setOutputStream(out);
Result result = new SAXResult(serializer.asContentHandler());

handler.setResult(result);
reader.parse(new InputSource(xmlInput));
FileDescriptor id = out.getFD();
in = new FileInputStream(id);
out.close();
xslInput.close();
xmlInput.close();

```

Figure 7.96 – Extrait du code source pour transformer le fichier XML intitulé « *xmlInput* » (nom de variable) en un fichier HTML « *out* » en fonction de la feuille de style « *xslInput* ».

Chaque feuille de style est associée à un processeur XSLT dédié à une transformation spécifique du document balisé. Nous décrivons dans la suite de cette partie les différentes feuilles de style associées à G-DEE permettant d'afficher le texte surligné, d'afficher le texte selon le modèle, et de visualiser des règles de décision dérivées.

7.3.3.2 Identification d'opérateurs déontiques dans le texte

Il est possible de surligner uniquement les opérateurs déontiques du texte analysé. Une feuille de style a été spécifiquement dédiée à cet affichage « à profondeur variable » et un extrait est représenté par les figures ci-dessous. La Figure 7.97 correspond à un en-tête typique d'un fichier XSL pour lequel on spécifie la structure du fichier généré suite à la transformation. Dans G-DEE, tous les fichiers générés par le processeur XSLT sont des fichiers au format HTML.

⁹⁷ <http://java.sun.com/j2se/1.4.2/docs/api/javax/xml/parsers/SAXParser.html> (dernier accès le 10 juin 2006).

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="html" encoding="ISO-8859-1"/>
<xsl:template match="/">
  <html>
    <head>
    </head>
    <body>
      <xsl:apply-templates/>
    </body>
  </html>
</xsl:template>
```

Figure 7.97 – En-tête du fichier XSL définissant la structure du fichier HTML après exécution du processeur XSLT.

La Figure 7.98 représente les règles suivies par le processeur XSLT, plus précisément, les balises `<xsl:template>` permettent de définir les actions à effectuer sur les balises concernées. Par exemple, le contenu de la balise *FrontScope* du document XML a une couleur de police de caractères noire pour le fichier HTML généré, ainsi que pour le contenu de la balise *BackScope*. Alors que le contenu de la balise *OpReco* a pour couleur de police de caractères le rouge, permettant ainsi de surligner les opérateurs déontiques correspondants.

```
<xsl:template match="FrontScope">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#000000;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="BackScope">
  <xsl:element name="font" >
    <xsl:attribute name="style">color:#000000;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="OpReco">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
    <b><xsl:value-of select="."/></b>
  </xsl:element>
</xsl:template>
</xsl:stylesheet>
```

Figure 7.98 – Description des templates pour l’affichage en rouge de l’opérateur déontique (*OpReco*).

7.3.3.3 Identification du *Front-scope* dans le texte

G-DEE permet également d’afficher sélectivement le *front-scope* et l’opérateur déontique. La feuille de style décrite ci-dessous correspond donc à ce type. La feuille de style contient toujours le même en-tête que celui décrit dans la Figure 7.97 mais le corps de la feuille de

style intègre la fonctionnalité de surlignage réservé au contenu du *FrontScope* en jaune comme décrit dans la figure ci-dessous par la propriété de l'attribut « *background-color: #FFFF66* ».

```
<xsl:template match="FrontScope">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#FFFF66;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="BackScope">
  <xsl:element name="font" >
    <xsl:attribute name="style">color:#000000;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="OpReco">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
    <b><xsl:value-of select="."/></b>
  </xsl:element>
</xsl:template>
</xsl:stylesheet>
```

Figure 7.99 – Description du fichier XSL pour surligner le contenu du *FrontScope* en jaune et le contenu de l'opérateur déontique en rouge (*OpReco*).

7.3.3.4 Identification du *Back-scope* dans le texte

La feuille de style décrite ci-dessous correspond à la fonctionnalité de surlignage sélectif du *back-scope* en bleu (avec l'opérateur déontique en police rouge).

```
<xsl:template match="FrontScope">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#000000;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="BackScope">
  <xsl:element name="font" >
    <xsl:attribute name="style">color:#99CCFF;</xsl:attribute>
    <xsl:value-of select="."/>
  </xsl:element>
</xsl:template>

<xsl:template match="OpReco">
  <xsl:element name="font">
    <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
    <b><xsl:value-of select="."/></b>
  </xsl:element>
</xsl:template>
</xsl:stylesheet>
```

Figure 7.100 – Description du fichier XSL pour surligner le contenu du *BackScope* en bleu et le contenu de l'opérateur déontique en rouge (*OpReco*).

7.3.3.5 Surlignement du *front-scope*, du *back-scope* et de l'opérateur déontique dans le texte

La fonctionnalité par défaut dans G-DEE consiste à surligner en jaune le *front-scope* et en bleu le *back-scope*. La figure ci-dessous représente l'en-tête du fichier XSL permettant de réaliser cette mise en forme. L'en-tête contient l'ensemble des variables nécessaires telles que la condition du *FrontScope* et la condition du *BackScope*, permettant d'isoler les marqueurs de condition dans la phrase balisée pour leur appliquer une mise en forme spécifique (en caractère gras).

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="html" encoding="ISO-8859-1"/>
<xsl:strip-space elements="texte"/>
<xsl:variable name="var_condition_fs">
  <xsl:value-of select="//FrontScope/condition"/>
</xsl:variable>
<xsl:variable name="var_condition_bs">
  <xsl:value-of select="//BackScope/condition"/>
</xsl:variable>
<xsl:variable name="var_fs">
  <xsl:value-of select="//FrontScope"/>
</xsl:variable>
<xsl:variable name="var_vide"> </xsl:variable>
  <xsl:template match="/">
    <html>
      <head>
      </head>
      <body>
        <xsl:apply-templates/>
      </body>
    </html>
  </xsl:template>
```

Figure 7.101 – Extrait de l'en-tête du fichier XSL définissant un ensemble de variables et spécifiant le fichier généré en HTML selon certaines règles définies dans le reste de la feuille de style.

La Figure 7.102 représente le reste du document XSL (Figure 7.101) qui définit les mises en formes à appliquer au contenu des balises. L'extrait ci-dessous correspond au traitement du contenu du *front-scope*, pour lequel deux mises en forme différentes sont possibles.

Soit la phrase balisée contient une balise indiquant une condition dans le *front-scope*, auquel cas la condition aura en plus du surlignement jaune une police de caractère en gras ; soit aucune balise indiquant une condition n'est présente dans le *front-scope*, et le contenu du *front-scope* est alors surligné en jaune.

```

<xsl:template match="FrontScope">
  <xsl:choose>
    <xsl:when test="$var_condition_fs=$var_vide">
      <xsl:element name="font">
        <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
        <xsl:value-of select="."/>
      </xsl:element>
    </xsl:when>
    <xsl:otherwise>
      <xsl:element name="font">
        <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
        <xsl:copy-of select="text()"/>
      </xsl:element>
      <xsl:if test="cond">
        <xsl:if test="true()">
          <xsl:element name="font">
            <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
            <xsl:value-of select="cond"/>
          </xsl:element>
        </xsl:if>
      </xsl:if>
      <xsl:if test="condition">
        <xsl:if test="true()">
          <xsl:element name="font">
            <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
            <b><xsl:value-of select="condition"/></b>
          </xsl:element>
        </xsl:if>
      </xsl:if>
    </xsl:otherwise>
  </xsl:choose>
  .....
</xsl:template>

```

Figure 7.102 – Mise en forme de la partie condition de la phrase.

7.3.3.6 Affichage du texte selon le modèle GEM

Une fonctionnalité de G-DEE concerne la transformation du texte en éléments GEM. Pour ce faire, une feuille de style est définie dans laquelle l'en-tête contient l'ensemble des variables nécessaires à la caractérisation de chaque segment de phrase en fonction des balises (Figure 7.103).

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="html" encoding="ISO-8859-1"/>
<xsl:strip-space elements="texte"/>
<xsl:variable name="var_condition_fs">
  <xsl:value-of select="//FrontScope/condition"/>
</xsl:variable>
<xsl:variable name="var_cond_fs">
  <xsl:value-of select="//FrontScope/cond"/>
</xsl:variable>
....
<xsl:variable name="var_condition_bs">
  <xsl:value-of select="//BackScope/condition"/>
</xsl:variable>
....
<xsl:variable name="var_op_reco">
  <xsl:value-of select="//OpReco_P"/>
</xsl:variable>
....

```

Figure 7.103 – En-tête du fichier permettant d'identifier automatiquement les conditions et les actions, et les afficher ensuite.

7.3. Intégration des fonctionnalités

De la même manière que pour les autres feuilles de style, la Figure 7.104 représente les règles de transformation du fichier XML en fichier HTML correspondant à l’affichage du contenu des balises (des *scopes* et de l’opérateur déontique) en fonction des balises GEM (<*decision.variable*> et <*action*>)⁹⁸.

Cet extrait correspond au traitement d’une phrase contenant un marqueur de condition dans le *front-scope*. La règle consiste à afficher la balise de GEM <*Decision.variable*> et d’associer le contenu de la balise du front-scope contenant la condition. Le processus reste à l’identique pour la balise <*Action*> de GEM.

La Figure 7.105 représente les règles d’affichage pour une phrase ne contenant aucun marqueur de condition. La variable de décision de GEM correspond au contenu du *front-scope* et l’action au contenu du *back-scope*⁹⁹.

```
<xsl:template match="/">
  <html>
    <head>
    </head>
    <body>
      <xsl:choose>
        <xsl:when test="$var_cond_fs">
          <xsl:if test="true()">
            <h3>&lt;Decision.variable&gt;
              <xsl:element name="font">
                <xsl:attribute name="style">background-color:#FFFF66;</
xsl:attribute>
                <b><xsl:apply-templates select="$var_cond_fs"/></b>
                <b><xsl:apply-templates select="$var_condition_fs"/></b>
              </xsl:element>
              &lt;/Decision.variable&gt;
            </h3>
            <h3>&lt;OpReco&gt;
              <xsl:element name="font">
                <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
                <b><xsl:apply-templates select="$var_op_reco"/></b>
              </xsl:element>
              &lt;/OpReco&gt;
            </h3>
            <xsl:choose>
              <xsl:when test="$var_subscope_fs=$var_virg">
                <xsl:choose>
                  ...
                </xsl:choose>
              </xsl:when>
              ....
            </xsl:when>
            <xsl:when test="$var_cond_bs">
              ....
            </xsl:when>
          </xsl:if>
        </xsl:when>
      </xsl:choose>
    </body>
  </html>
</xsl:template>
```

Figure 7.104 – Extrait du fichier XSL correspondant à une phrase contenant une condition dans le *front-scope*.

⁹⁸ Dans un souci de conserver tous les segments textuels, nous avons ajouté une balise <OpReco> dans le modèle GEM permettant d’isoler l’opérateur déontique.

⁹⁹ Cette règle est définie pour les phrases actives. G-DEE contient un module permettant de dissocier une phrase à la voix active et passive ce qui permet d’activer à ce stade du traitement la feuille de style appropriée.

```

.....
<xsl:otherwise>
  <h3>&lt;Decision.variable&gt;
    <xsl:element name="font">
      <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
      <xsl:apply-templates select="$var_dv_fs"/>
    </xsl:element>
    &lt;/Decision.variable&gt;
  </h3>
  <h3>&lt;OpReco&gt;
    <xsl:element name="font">
      <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
      <b><xsl:apply-templates select="$var_op_reco"/></b>
    </xsl:element>
    &lt;/OpReco&gt;
  </h3>
  <h3>&lt;Action&gt;
    <xsl:element name="font">
      <xsl:attribute name="style">background-color:#99CCFF;</xsl:attribute>
      <b><xsl:apply-templates select="$var_action_bs"/></b>
    </xsl:element>
    &lt;/Action&gt;
  </h3>
</xsl:otherwise>

```

Figure 7.105 – Affichage des éléments GEM pour une phrase ne contenant pas de marqueurs de condition.

7.3.3.7 Affichage du texte selon une règle de décision

L'en-tête est identique à la Figure 7.103 pour ce qui concerne la définition des variables nécessaires à la transformation du texte sous la forme d'une règle de décision. La figure ci-dessous représente la règle applicable pour une phrase contenant une condition dans le *back-scope*.

```

<xsl:when test="$var_cond_bs">
  <xsl:if test="true()">
    <h3>SI
      <xsl:element name="font">
        <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>
        <b><xsl:apply-templates select="$var_condition_bs"/></b>
      </xsl:element>
    </h3>
    <h3>ALORS
      <xsl:choose>
        <xsl:when test="$var_op_reco_inf">
          <xsl:element name="font">
            <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
            <b><xsl:apply-templates select="$var_op_reco_inf"/></b>
          </xsl:element>
        </xsl:when>
        <xsl:otherwise>
          <xsl:element name="font">
            <xsl:attribute name="style">color:#FF0000;</xsl:attribute>
            <b><xsl:apply-templates select="$var_op_reco"/></b>
          </xsl:element>
        </xsl:otherwise>
      </xsl:choose>
    </h3>
  </xsl:if>
  .....

```

Figure 7.106 – Traitement d'une phrase contenant une condition dans le *back-scope*.

7.3. Intégration des fonctionnalités

La partie « SI » de la règle (précédée dans la feuille de style de la balise `<h3>` pour l’affichage du fichier HTML généré) reprend le contenu du *back-scope* et la partie « ALORS » recouvre le contenu de la balise de l’opérateur déontique et celui du *front-scope*.

La Figure 7.107 concerne l’extrait de la feuille de style XSL correspondant au traitement d’une phrase ne contenant pas de marqueur de condition. La partie « SI » de la règle contient le contenu du *front-scope* et la partie « ALORS » l’opérateur déontique et le *back-scope*¹⁰⁰.

```
<xsl:otherwise>
  <h3>SI
    <xsl:element name="font">
      <xsl:attribute name="style">background-color:#FFFF66;</xsl:attribute>

      <xsl:apply-templates select="$var_dv_fs"/>
    </xsl:element>
  </h3>
  <h3>ALORS
    <xsl:choose>
      <xsl:when test="$var_op_reco_inf">
        <xsl:element name="font">
          <xsl:attribute name="style">color:#FF0000;</xsl:attribute>

          <b><xsl:apply-templates select="$var_op_reco_inf"/></b>
        </xsl:element>
      </xsl:when>
      <xsl:otherwise>
        <xsl:element name="font">
          <xsl:attribute name="style">color:#FF0000;</xsl:attribute>

          <b><xsl:apply-templates select="$var_op_reco"/></b>
        </xsl:element>
      </xsl:otherwise>
    </xsl:choose>
  </h3>
  <h3>
    <xsl:element name="font">
      <xsl:attribute name="style">background-color:#99CCFF;</xsl:attribute>
      <b><xsl:apply-templates select="$var_action_bs"/></b>
    </xsl:element>
  </h3>
</xsl:otherwise>
```

Figure 7.107 – Extrait de la feuille de style XSL correspondant au traitement d’une phrase ne contenant pas de condition.

Le processeur XSLT est une technique efficace et rapide pour représenter les informations voulues, selon un affichage personnalisé comme la représentation selon le modèle GEM ou de règles de décision. Néanmoins, une condition nécessaire à ce type de traitement est de disposer d’un balisage spécifique de la phrase, permettant de reconnaître chaque segment textuel durant le traitement.

¹⁰⁰ De la même manière que pour les autres feuilles de style, cet extrait concerne le cas d’une phrase à la voix active. Une autre feuille de style correspond au traitement des phrases à la voix passive.

Chapitre 8

Couverture syntaxique et traitement de formes particulières

Ce chapitre présente certains phénomènes syntaxiques « réels » rencontrés dans les Guides de Bonnes Pratiques et qui sont traités par G-DEE, et est donc l'occasion de confronter le système aux limites de sa performance. Ces différents exemples de phrases représentant des phénomènes linguistiques spécifiques permettront aussi, au-delà de certaines limitations du système, de mettre en évidence certains problèmes de qualité d'écriture des Guides de Bonnes Pratiques qui se manifestent par des structures atypiques.

8.1 Préambule

Ce chapitre présente les résultats de l'analyseur en illustrant plus particulièrement certains phénomènes linguistiques rencontrés dans les Guides de Bonnes Pratiques. Ces phénomènes font intervenir des règles spécifiques de traitement qui sont détaillées pour chaque exemple, et concernent soit le pré-traitement soit les motifs syntaxiques. Il s'agit ici de montrer a posteriori que les descriptions linguistiques et les techniques associées sont bien un aspect indissociable du balisage intelligent ou de la structuration semi-automatique des Guides de Bonnes Pratiques. En d'autres termes, des techniques empiriques de recherche d'informations ou de manipulation de mots-clés ne peuvent suffire à implémenter des fonctionnalités avancées de structuration des documents. Un aperçu des phénomènes linguistiques traités permet aussi de discuter de la complexité du développement de G-DEE et de sa portabilité à d'autres domaines de spécialité.

Les exemples ci-dessous représentent des phrases balisées par l'analyseur de G-DEE, pour lesquelles l'opérateur déontique est souligné, le *front-scope* est en italique et le *back-scope* en caractères gras.

8.2 Pré-traitement et « phrases réelles »

S'il est un fait bien établi en Traitement Automatique des Langues (TAL), c'est bien la distance qui sépare les « exemples linguistiques » phrases isolées qui illustrent un problème syntaxique et/ou sémantique particulier des phrases extraites de textes réels, issus de corpus attestés. Les systèmes d'Extraction d'Informations étant par nature confrontés à ces dernières, leurs concepteurs ont vite réalisé l'inadéquation d'analyseurs traditionnels, même les plus empiriques, face à la complexité et parfois à l'hétérogénéité des phrases issues de ces textes. Ces systèmes ont donc développé des approches empiriques du pré-traitement et de la segmentation des phrases qui sont vite devenues partie intégrante de l'Extraction d'Informations.

En ce qui concerne nos propres traitements, nous avons naturellement formulé un certain nombre d'hypothèses de départ qui souffraient des travers traditionnels du TAL. Nous avons au départ étudié des recommandations caractéristiques qui nous ont dans un premier temps conduit à poser des hypothèses simplificatrices, comme le fait qu'une phrase contiendrait une recommandation unique, signalée par un seul opérateur déontique. Cette hypothèse ignorait une réalité plus complexe : d'une part certaines phrases, à la suite d'un processus d'écriture complexe et non exempt de compromis, peuvent avoir des structures non-univoques ; d'autre part tous les verbes déontiques (pouvoir, devoir ...) ne signalent pas un opérateur déontique et il peut exister des cas de « faux positifs ». Tout ceci concourt à faire en sorte qu'une même phrase puisse contenir deux opérateurs déontiques (vrais ou supposés du fait d'un faux positif). Une fois ce phénomène reconnu, il apparaît que son traitement, tant pour des raisons théoriques que pratiques, relève davantage de tâches de

8.1. Préambule

pré-traitement que d'une complexification des techniques d'analyse linguistique destinées à la syntaxe des opérateurs.

Les trois recommandations ci-dessous représentent des phrases nécessitant ce type de pré-traitement, du fait que chacune d'entre elles contient deux opérateurs déontiques. Le chapitre 6 a décrit en détail les étapes de pré-traitement pour ce type de configuration permettant de reconnaître deux opérateurs déontiques dans une phrase. Le pré-traitement, pour ces trois phrases, consiste donc à segmenter la phrase en fonction de la ponctuation.

Les deux exemples ci-dessous illustrent la reconnaissance de deux opérateurs déontiques à la voix passive. Le groupe verbal déontique est constitué de l'auxiliaire *être* et du verbe déontique au participe passé. La segmentation est réalisée simplement après la virgule.

- ***Une bithérapie est conseillée lorsque l'HbA1c est > 6,5% sur deux contrôles successifs effectués à 3-4 mois d'intervalle sous monothérapie orale à posologie optimale, elle est recommandée lorsque l'HbA1c est > 8%, sous monothérapie orale à posologie optimale (accord professionnel).***

Le traitement des deux propositions issues de la segmentation ne présente pas de problèmes particuliers du fait que le nombre d'occurrences de l'auxiliaire *être* est identique au nombre de verbes au participe passé de la phrase (ce qui ne serait pas le cas dans une forme « zeugmatique »). L'anaphore pronominale n'est pas traitée dans la version actuelle du prototype, cependant il apparaît que l'adjonction d'heuristiques simples (mémorisation des référents sur une pile) suffirait à couvrir une grande proportion des cas observés.

- ***L'utilisation d'échelles du type suivant est recommandée pour l'état neurologique, plusieurs échelles spécifiques à l'AVC sont utilisées, en langue française ou anglaise ou en traduction française : échelle d'Orgogozo, échelle canadienne, échelle scandinave, échelle du National Institutes of Health (NIH).***

Les groupes verbaux déontiques de l'exemple suivant sont composés de verbes déontiques à la voix passive.

- ***L'association statine + fénofibrate est classiquement déconseillée ; elle peut se discuter après avis spécialisé.***

Nous pouvons relever la pertinence ou non de deux opérateurs déontiques dans une recommandation. En général, comme dans l'exemple ci-dessus, l'association d'un opérateur déontique à un verbe (*discuter*) correspond à des conditions d'applications d'une démarche. Elles ont donc une signification pratique : si les deux opérateurs sont confondus on obtient une recommandation approximative qui perd la nuance initiale, mais la lecture du texte du *back-scope* permet en général de retrouver cette nuance. On peut imaginer comme une extension du système la détection automatique de ces recommandations « conditionnelles ».

8.3 Formes adjectivales

Nous avons assimilé les opérateurs déontiques à des verbes, néanmoins il existe aussi des formes adjectivales qui doivent être considérées comme des opérateurs déontiques à part entière, et qui sont susceptibles de structurer de la même façon une recommandation autour d'eux, le *front-scope* étant constitué par le nom prédiqué par l'adjectif déontique.

Les deux exemples ci-dessous représentent l'analyse de ce type de syntaxe souvent employé dans les Guides de Bonnes Pratiques. Les formes déontiques adjectivales posent un problème de redéfinition (et, au plan de l'analyse, de délimitation) de l'opérateur déontique. Les motifs définis considèrent l'opérateur comme délimité par l'adjectif déontique et la première occurrence verbale, le verbe *être* dans les exemples qui suivent :

- *Les examens complémentaires recommandés dans le cadre du bilan initial sont :*
 - **créatininémie et estimation du débit de filtration glomérulaire (DFG : formule de Cockcroft et Gault) ;**
 - **bandelette réactive urinaire (protéinurie, hématurie) et quantification si positivité ;**
 - **kaliémie (sans garrot) ;**
 - **prélèvements à jeun : glycémie ;**
 - **cholestérol total et HDL-cholestérol, triglycérides ;**
 - **calcul du LDL-cholestérol (formule de Friedewal) ;**
 - **ECG de repos.**

- *En dehors des cas particuliers, les classes thérapeutiques recommandées en première intention chez le patient âgé sont les diurétiques thiazidiques ou les inhibiteurs calciques de type dihydropyridine de longue durée d'action (Grade A).*

La rédaction des Guides de Bonnes Pratiques ne suit pas de consignes de style particulières ce qui explique certaines formes plus complexes et indirectes, dont voici deux illustrations :

- *Lorsqu'un patient est examiné alors que le déficit neurologique persiste, l'attitude médicale recommandée est de considérer que ce patient est atteint d'un AVC aigu, dont les symptômes peuvent ou non régresser rapidement, et de mettre en œuvre, en urgence les moyens diagnostiques et thérapeutiques appropriés.*

- *Lorsqu'un patient est examiné alors que le déficit neurologique a régressé, l'attitude médicale recommandée est de mettre en œuvre dans les meilleurs délais les moyens diagnostiques et thérapeutiques détaillés ci-dessous dans le but d'initier au plus vite un traitement de prévention secondaire.*

Les mêmes motifs définis plus haut sont cependant capables de détecter les opérateurs déontiques de ce type.

8.4 Formes particulières utilisant les auxiliaires et les verbes déontiques

Les verbes déontiques sont d'une grande utilisation en français y compris dans des locutions courantes, et ceci peut être une source de confusion pour la reconnaissance d'opérateurs déontiques. D'une façon analogue, des auxiliaires ou des verbes supports peuvent s'insérer dans des expressions déontiques.

Une règle de traitement consiste à identifier le positionnement du verbe déontique générique. Dans le contexte des recommandations, un verbe déontique générique à la 3^{ème} personne du singulier ou du pluriel ne peut pas être positionné en fin de phrase. Cette règle générique rend la reconnaissance robuste à l'insertion de locutions en fin de phrase (« dès que l'on pourra », « autant que faire se peut », etc.).

- *Le lever précoce est recommandé autant que faire se peut.*

Les deux exemples suivants représentent des recommandations basées sur les verbe *éviter* et *utiliser* mais pour lesquelles le verbe *être* se comporte en verbe support.

- *Les voies intramusculaire et sublinguale sont à éviter.*
- *La rosuvastatine est à utiliser en cas d'intolérance ou d'efficacité insuffisante des autres statines.*

D'autres phrases ne sont pas basées principalement sur des opérateurs déontiques mais impliquent potentiellement une recommandation. Pour détecter ce type de syntaxe, un motif syntaxique a été ajouté : il prend comme point de départ une forme impersonnelle signalant l'existence (« il existe », « il y a » ...) suivi d'une préposition et d'un verbe, généralement un opérateur déontique à l'infinitif. La couverture apportée par ce motif permet de reconnaître avec succès les exemples suivants.

- *Il n'existe pas de preuves suffisantes pour recommander, en routine, la contention des membres inférieurs en prévention des complications thrombo-emboliques veineuses chez les patients ayant un AVC ischémique.*
- *Ultérieurement, il n'existe pas de données scientifiques permettant de proposer une périodicité à ces contrôles.*
- *Il n'y a pas de justification scientifique à pratiquer un dosage initial systématique des CPK dans la population générale.*
- *Il n'y a pas d'arguments pour proposer un traitement au long cours.*

- *Si le taux de CA est initialement élevé, il n'y a pas lieu de doser d'autres marqueurs en routine (standard, accord d'experts).*

Les verbes ou expressions déontiques peuvent être introduits par des expressions variables et c'est pourquoi nous avons introduit un motif approprié pour les verbes support introduisant des verbes déontiques à l'infinitif comme dans les exemples suivants :

- *Chez les patients dont la PA est comprise entre 40-179/90-109 mm Hg et qui n'ont pas un risque élevé, il est nécessaire de confirmer cette HTA au moyen de l'automesure ou de la MAPA.*
- *Chez le sujet âgé et chez le sujet athéroscléreux, lorsqu'un examen échographique a été prescrit pour un autre motif que l'HTA, il est souhaitable de profiter de cet examen pour vérifier l'absence d'anévrisme de l'aorte compte tenu de la prévalence élevée de cette lésion chez le sujet âgé.*
- *Il n'est pas justifié de répéter le bilan, sauf en cas d'apparition d'un facteur de risque cardiovasculaire.*
- *Ainsi, il convient d'éviter les régimes trop restrictifs conduisant à des déséquilibres alimentaires et à des troubles du comportement alimentaire.*
- *Concernant les risques hépatiques et musculaires liés aux statines, il convient de prendre en compte les éléments suivants.*

Chapitre 9

Performances et évaluation préliminaire de G-DEE

Ce chapitre présente deux expériences d'évaluation de G-DEE. Ce chapitre intéressera les personnes s'attachant aux performances techniques de G-DEE en tant que base d'un outil opérationnel dans le futur, ainsi que les impacts observés auprès d'experts médicaux. La première expérience que nous avons menée consiste à évaluer les performances techniques de G-DEE définies par le taux de reconnaissance des recommandations en comparaison à un balisage manuel d'un ensemble de Guides de Bonnes Pratiques. La deuxième expérience consiste à faire évaluer les Guides de Bonnes Pratiques par des experts intervenant tant au niveau de l'élaboration que de l'évaluation des Guides de Bonnes Pratiques, et impliqués dans la mise en œuvre de ces Guides. Une grille de cotations, inspirée par les évaluations publiées pour les systèmes d'Extraction d'Informations ou de Recherche d'Informations, a été définie. Les résultats seront analysés en ce qui concerne la validité de l'approche et permettront de poser différentes problématiques quant à l'écriture et l'élaboration de ces documents textuels, qui peuvent constituer autant de pistes méthodologiques pour l'amélioration du cycle de vie des Guides de Bonnes Pratiques.

9.1 Évaluation préliminaire par un jeu de test

Cette première expérience d'évaluation étudie la performance de G-DEE sur sa tâche élémentaire d'identification de recommandations élémentaires dans un texte. Le principe de l'évaluation consiste à comparer le balisage d'opérateurs déontiques par G-DEE dans des textes (sans intersection avec les corpus de référence) à un balisage manuel de ces mêmes textes qui servira de référence.

Le jeu de test que nous avons créé est constitué de cinq Guides de Bonnes Pratiques (1 003 phrases – 83 997 mots) élaborés par l'HAS¹⁰¹ (Haute Autorité de Santé) qui porte sur les thèmes du diabète (correspondant à une actualisation récente¹⁰²), des accidents ischémiques transitoires¹⁰³, de l'hypertrophie de la prostate¹⁰⁴, de l'accident vasculaire cérébral¹⁰⁵, et du carcinome basocellulaire¹⁰⁶. Ces cinq Guides de Bonnes Pratiques n'ont pas été utilisés pour définir notre grammaire d'opérateurs déontiques.

Le jeu de test est dans un premier temps balisé manuellement et seuls les opérateurs déontiques basés sur les verbes (déontiques) *recommander*, *devoir* et *pouvoir* sont recherchés. Au total, 311 occurrences d'opérateurs déontiques ont été reconnues dans le jeu de test. Le fait de limiter le balisage à ces trois verbes déontiques est justifié par leur représentativité suffisante des textes des Guides de Bonnes Pratiques (comme décrit dans le chapitre 3).

Dans le but d'évaluer la performance de G-DEE nous comparons les phrases balisées manuellement sur ce jeu de test avec celles balisées automatiquement par G-DEE. Un balisage est considéré comme « correct » lorsque l'opérateur déontique est reconnu et balisé par G-DEE, ainsi que ses *scopes* associés. Pour chaque Guide de Bonnes Pratiques, nous comptons simplement le nombre d'occurrences des verbes déontiques correctement balisés par le système.

A l'issue de la comparaison des recommandations balisées manuellement et celles balisées automatiquement par G-DEE, nous obtenons un score de 97% d'occurrences d'opérateurs déontiques correctement balisés par G-DEE, sur un ensemble de 1 003 phrases. Les 3% d'erreurs identifiées proviennent de phénomènes syntaxiques spécifiques suggérant qu'une extension de la couverture de reconnaissance de la grammaire peut améliorer le taux de reconnaissance pour ce jeu de test. Par exemple, la négation suivie d'une conjonction pour l'Extrait 1 et la forme adjectivale du verbe *recommander* pour l'Extrait 2.

¹⁰¹ <http://ww.has-sante.fr/> (dernier accès le 10 juin 2006).

¹⁰² <http://afssaps.sante.fr/> (en cours de publication).

¹⁰³ Prise en charge diagnostique et traitement immédiat de l'accident ischémique transitoire de l'adulte (Mai 2004).¹⁰¹

¹⁰⁴ Prise en charge diagnostique et thérapeutique de l'hypertrophie bénigne de la prostate (Mars 2003).¹⁰¹

¹⁰⁵ Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral – Aspects médicaux (Septembre 2002).¹⁰¹

¹⁰⁶ Prise en charge diagnostique et thérapeutique du carcinome basocellulaire de l'adulte (Mars 2004).¹⁰¹

Extrait 1

La radiothérapie n'est recommandée qu'en 2e intention si la chirurgie n'est pas possible.

Extrait 2

Lorsqu'un patient est examiné alors que le déficit neurologique a régressé, l'attitude médicale recommandée est de mettre en œuvre dans les meilleurs délais les moyens diagnostiques et thérapeutiques détaillés ci-dessous dans le but d'initier au plus vite un traitement de prévention secondaire.

9.2 Évaluation par des experts

L'évaluation ci-dessus mesure une performance « brute » du système et s'intéresse aux fonctionnalités techniques de base du système. Elle constitue une première étape importante en particulier pour déterminer le potentiel d'amélioration de la couverture des composants d'analyse linguistique.

Une évaluation plus réaliste se doit de prendre en compte le contexte et la finalité des Guides de Bonnes Pratiques. Le but de cette seconde expérience consiste à faire évaluer par des experts deux Guides de Bonnes Pratiques balisés par G-DEE, plus précisément de vérifier le balisage généré par G-DEE en utilisant une grille de cotations empruntée aux techniques d'évaluation de certains systèmes d'Extraction d'Informations. Ces grilles cherchent à comparer la reconnaissance d'informations (ici, de recommandations) par le système à celle d'experts qui procèdent naturellement à une analyse plus fine, et pas toujours exemptes de critiques, du contenu des Guides de Bonnes Pratiques. Les experts sélectionnés ont à la fois une connaissance du domaine et une pratique des Guides de Bonnes Pratiques en particulier de leur rédaction : ceci leur permet de formuler un avis quant à la caractérisation ou non d'une phrase en tant que recommandation, quelle que soit sa forme de surface.

Les Guides de Bonnes Pratiques servant de support à cette expérience sont la « Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral. » (ANAES, 2002)¹⁰⁷ et la « Prise en charge des patients adultes atteints d'hypertension artérielle essentielle – Actualisation 2005 » (HAS, 2005)¹⁰¹.

Cette expérience fait intervenir trois experts évaluant une à une les 103 phrases du Guide de Bonnes Pratiques de l'accident vasculaire cérébral (AVC), puis quatre experts évaluant les 142 phrases du Guide de Bonnes Pratiques de l'hypertension artérielle (HTA). Leurs cotations sont ensuite comparées aux résultats produits par G-DEE en utilisant des métriques telles que la précision, la spécificité et la sensibilité (Tableau 9.12 et Tableau 9.14). Enfin, les résultats sont discutés avec chaque expert, et certains conflits potentiels sont interprétés avec l'ensemble des experts.

¹⁰⁷ Le Guide de Bonnes Pratiques de l'AVC analysé par G-DEE est présenté dans l'annexe 4.

9.2.1 Protocole d'évaluation

Le travail de chaque expert consiste à vérifier que chaque phrase est correctement balisée et qu'elle correspond ou non à une recommandation. La grille de cotations fournie aux experts est composée des critères d'évaluation suivants : (i) les phrases correctement balisées (vraies positives VP) correspondent à une recommandation effective pour laquelle l'opérateur déontique est marqué en rouge gras ; le *front-scope* est marqué en jaune ; et le *back-scope* est marqué en bleu. Par exemple,

Les corticostéroïdes **ne doivent pas être utilisées** en cas d'œdème cérébral (grade B), car ils sont inefficaces et dangereux.

(ii) les phrases fausses positives (FP) correspondent à des phrases balisées qui ne devraient pas l'être car elles ne correspondent pas à des recommandations. Par exemple,

L'échelle du NIH **est** l'échelle de référence en cas de fibrinolyse.

(iii) les phrases fausses négatives (FN) correspondent à des recommandations qui n'ont pas été ou qui ont été incorrectement balisées. L'exemple décrit ci-dessous illustre le non balisage de l'opérateur déontique « *il est recommandé* »,

Dans le cas d'une HTA essentielle (hormis en cas de diabète ou d'atteinte rénale associé) il est recommandé de faire baisser les chiffres tensionnels en dessous de 140 mm Hg pour la PAS et 90 mm Hg pour la PAD.

Nous avons opté pour une évaluation qui préfère sous-estimer la performance, en n'attribuant pas de score partiel, mais un score de zéro (FN), pour les recommandations incomplètement balisées.

Et (iv) les phrases vraies négatives (VN) correspondent à des phrases qui n'ont pas été balisées et qui ne doivent pas être balisées. Par exemple, la phrase ci-dessous correspond à un constat.

L'HTA systolique est plus fréquente chez le sujet âgé.

9.2.2 Étude de la concordance entre experts

Pour cette évaluation, le coefficient du Kappa (Grenier, 1990, 1996) est calculé compte tenu du fait qu'il est considéré comme un outil simple pour mesurer la concordance entre deux jugements à plusieurs catégories. Le Tableau 9.7 représente le nombre de phrases concordantes entre deux experts pour le Guide de Bonnes Pratiques de l'AVC.

9.2. Evaluation par des experts

Tableau 9.7 – Nombre de phrases concordantes entre deux experts pour la mesure du Kappa (Guide de Bonnes Pratiques de l'AVC).

		Expert 1			
		VP	FP	FN	VN
Expert 2	VP	64	1	0	0
	FP	1	2	0	0
	FN	1	0	0	4
	VN	0	0	1	41

Nous avons calculé les coefficients du Kappa entre les experts (Tableau 9.8), et en suivant la définition de l'interprétation des valeurs du coefficient du Kappa, elle est excellente entre l'expert 1 et l'expert 2, et bonne pour le reste des experts.

Tableau 9.8 – Coefficient du Kappa pour le Guide de Bonnes Pratiques de l'AVC.

	Expert 1	Expert 2	Expert 3
Expert 1		0,87	0,70
Expert 2			0,71
Expert 3			

De la même manière le tableau ci-dessous représente le nombre de phrases concordantes entre deux experts pour le Guide de Bonnes Pratiques de l'HTA.

Tableau 9.9 – Nombre de phrases concordantes entre deux experts pour le Guide de Bonnes Pratiques de l'HTA.

		Expert 1			
		VP	FP	FN	VN
Expert 4	VP	77	4	0	0
	FP	0	8	0	0
	FN	0	0	7	6
	VN	0	0	1	50

Les coefficients du Kappa sont également calculés pour mesurer la concordance entre les experts pour ce Guide et sont représentés par le tableau ci-dessous.

Tableau 9.10 – Coefficient du Kappa pour le Guide de Bonnes Pratiques de l'HTA.

	Expert 1	Expert 2	Expert 3	Expert 4
Expert 1		0,71	0,74	0,75
Expert 2			0,74	0,88
Expert 3				0,77
Expert 4				

On peut observer une concordance excellente entre l'expert 2 et 4 (coefficient du Kappa de 0,88), et pour le reste des experts, la concordance reste bonne.

9.2.3 Étude des accords entre experts

Les accords forts entre experts sont une mesure simple d'améliorer la couverture syntaxique de G-DEE. Pour l'Extrait 3, par exemple, les trois experts considèrent que la phrase n'est pas une recommandation. Les justifications fournies sont « Ce n'est pas une recommandation, c'est un constat », « Ce n'est pas une recommandation », et « c'est un problème de l'argumentation négative ».

Extrait 3

« Les indications du traitement neurochirurgical des AVC ischémiques ou hémorragiques sont insuffisamment évaluées. »

Les quatre experts ont également considéré que l'Extrait 4 qui n'était pas balisé par G-DEE relevait du même esprit que l'Extrait 5 balisé par G-DEE (et validé par les quatre experts comme une recommandation).

Extrait 4

« Chez le patient diabétique, l'objectif tensionnel est une PA < 130/80 mm Hg. »

Extrait 5

« Dans le cas d'une HTA essentielle (hormis en cas de diabète ou d'atteinte rénale) il est recommandé de faire baisser les chiffres tensionnels en dessous de 140 mm Hg pour la PAS et 90 mm Hg pour la PAD. »

9.2.4 Étude des désaccords entre experts

Chaque expert a évalué les phrases des Guides de Bonnes Pratiques en contexte, ce qui n'est pas sans conséquences quant aux jugements qu'ils ont formulés. Il n'est pas inutile de rappeler que la rédaction des Guides de Bonnes Pratiques n'est normalisée ni sur la forme ni sur le contenu. En conséquence, il peut y avoir des désaccords entre experts sur le fond (une phrase correspond à une attitude clinique recommandée ou non, donc à une recommandation potentielle) et/ou sur la forme (une phrase peut être une recommandation même en l'absence de marqueurs de surface, opérateurs déontiques ou autres). Les experts ont été consultés pour justifier les FN et FP qu'ils ont identifiés. Nous présentons également les justifications orales (en Annexe 5) qu'ils ont pu fournir permettant de comprendre la difficulté rencontrée pour l'évaluation de certaines phrases.

9.2.5 Performance « brute » de G-DEE

Nous décomptons pour chaque expert et chaque Guide de Bonnes Pratiques le nombre de VP, VN, FP et FN. Puis, pour interpréter ces résultats nous utilisons les métriques traditionnelles (van Rijsbergen, 1979) suivantes :

- la précision correspondant au pourcentage de phrases balisées qui sont effectivement des recommandations [VP / (VP + FP)]

9.2. Evaluation par des experts

- le rappel (sensibilité) correspondant au pourcentage de phrases qui sont des recommandations et qui sont bien balisées comme telles [VP / (VP + FN)].
- le bruit correspondant au pourcentage de phrases balisées qui ne devraient pas l'être [FP / (FP + VP)].
- la spécificité correspondant aux phrases non balisées et qui ne sont pas des recommandations [VN / (FP + VN)].
- la F-mesure : correspondant à la combinaison du rappel et de la précision (Hripcsak & Rothschild, 2005) [$F = ((1 + \beta^2) * \text{rappel} * \text{précision}) / ((\beta^2 * \text{précision}) + \text{rappel})$].

Les tableaux ci-dessous représentent les résultats obtenus pour le Guide de Bonnes Pratiques de l'AVC. On observe que le pourcentage de phrases qui correspondent à des recommandations et qui ont été correctement balisées par G-DEE varie de façon significative selon les experts [81 – 99%].

Tableau 9.11 – Résultats détaillés de l'évaluation pour le Guide de Bonnes Pratiques de l'AVC.

	Nombre de phrases	Phrases considérées comme « recommandations » par les experts	Phrases considérées comme « non recommandations » par les experts
PHRASES BALISÉES PAR G-DEE			
Expert 1	74	68	6
Expert 2	70	67	3
Expert 3	77	71	7
PHRASES NON BALISÉES PAR G-DEE			
Expert 1	48	1	47
Expert 2	46	5	41
Expert 3	45	17	28

Tableau 9.12 – Résultats de l'évaluation de G-DEE pour le Guide de Bonnes Pratiques de l'AVC.

	Expert 1	Expert 2	Expert 3
Sensibilité	0,99 (IC 95% : 0,96 ; 1)	0,93 (IC 95% : 0,88 ; 0,98)	0,81 (IC 95% : 0,74 ; 0,88)
Spécificité	0,89 (IC 95% : 0,83 ; 0,94)	0,93 (IC 95% : 0,89 ; 0,98)	0,82 (IC 95% : 0,76 ; 0,89)
Précision	0,92	0,96	0,92
Bruit	0,11	0,07	0,18
F-mesure	0,95	0,94	0,86

L'évaluation montre aussi que peu de FP sont décomptés [3 – 6], mais on remarque surtout l'existence d'un désaccord pour les FN [1 - 17] correspondant aux recommandations qui n'ont pas été balisées ou incorrectement balisées. La F-mesure est bonne pour deux experts sur trois.

Les tableaux ci-dessous représentent de la même manière les résultats obtenus pour le Guide de Bonnes Pratiques de l'HTA. Le pourcentage de phrases qui correspondent à des recommandations et qui ont été correctement balisées par G-DEE est également variable selon les experts [83 – 91%].

Tableau 9.13 – Résultats détaillés de l'évaluation pour le Guide de Bonnes Pratiques de l'HTA.

	Nombre de phrases	Phrases considérées comme « recommandations » par les experts	Phrases considérées comme « non recommandations » par les experts
PHRASES BALISÉES PAR G-DEE			
Expert 1	88	77	11
Expert 2	88	84	4
Expert 3	88	86	2
Expert 4	88	81	7
PHRASES NON BALISÉES PAR G-DEE			
Expert 1	65	8	57
Expert 2	66	14	52
Expert 3	63	18	45
Expert 4	66	14	52

Tableau 9.14 – Résultats de l'évaluation de G-DEE pour le Guide de Bonnes Pratiques de l'HTA.

	Expert 1	Expert 2	Expert 3	Expert 4
Sensibilité	0,91 (IC 95% : 0,86 ; 0,95)	0,86 (IC 95% : 0,80 ; 0,91)	0,83 (IC 95% : 0,77 ; 0,89)	0,85 (IC 95% : 0,80 ; 0,91)
Spécificité	0,84 (IC 95% : 0,78 ; 0,90)	0,93 (IC 95% : 0,89 ; 0,97)	0,96 (IC 95% : 0,93 ; 0,99)	0,88 (IC 95% : 0,83 ; 0,93)
Précision	0,88	0,95	0,98	0,92
Bruit	0,16	0,07	0,04	0,12
F-mesure	0,89	0,90	0,90	0,89

On observe également peu de FP [2 – 11] pour ce Guide, mais on remarque encore des désaccords entre experts pour les FN [8 – 18]. Ce sont ces désaccords ainsi que leur mécanisme que nous allons explorer.

9.2.6 Les FN détectées par des experts

Voici un exemple de FN sujet à des désaccords entre les experts. Ces phrases correspondent à des phrases non balisées par G-DEE mais considérées comme des recommandations par les experts.

Par exemple, pour l'Extrait 6, un des experts considère que la phrase est une recommandation en la justifiant par « c'est une recommandation » alors que les deux autres experts l'ont évalué comme un VN.

9.2. Evaluation par des experts

Extrait 6

« Une surveillance continue par cardioscope au cours des 48 premières heures de l'AVC permet de dépister une arythmie paroxystique ou une autre pathologie cardiaque associée. »

L'extrait ci-dessous représente une phrase considérée par deux experts sur trois comme une recommandation.

Extrait 7

« En l'absence d'alimentation orale, un apport calorique adéquat sera fourni par voie entérale, à partir de la 48^e heure (accord professionnel). »

Un des experts justifie le fait que ce soit une recommandation en reformulant la phrase « C'est comme si on disait : il faut fournir ... », alors qu'un autre expert invoque directement une notion grammaticale (« c'est un futur impératif »), qui impliquerait une recommandation. Néanmoins, le troisième expert ne partage pas ce point de vue.

L'Extrait 8 représente une phrase considérée par un seul expert sur les trois comme étant une recommandation en invoquant sa connaissance du contexte : « c'est une recommandation générale de Santé Publique. Une recommandation pour les institutions et pour stimuler les cliniques. », qui constitue presque un argument « sociologique ». Les deux autres experts la considèrent comme un VN. On notera pourtant le caractère particulier de cette formulation qui, en donnant la parole au groupe de travail, rompt avec les conventions traditionnellement observées dans les Guides de Bonnes Pratiques pour se placer en quelque sorte à un niveau de « méta-recommandations ».

Extrait 8

« Le groupe considère qu'il est urgent d'offrir aux patients atteints d'AVC une prise en charge structurée intégrant les soins à la phase aiguë et les soins de suite et de réadaptation. »

Trois experts sur quatre considèrent que l'Extrait 9 correspond à une recommandation par la justification que le futur impératif implique une recommandation.

Extrait 9

« La baisse de la pression artérielle chez le patient âgé hypertendu de plus de 65 ans, sera progressive. »

Pour l'Extrait 10, deux experts sur quatre identifient la phrase comme une recommandation dite « d'objectif ».

Extrait 10

« En pratique au-delà de 80 ans, l'objectif thérapeutique est d'atteindre une PAS < 150 mm Hg, en l'absence d'hypotension orthostatique (Grade B). »

Ces extraits nous montrent une variabilité d'interprétation de ce qu'un expert considère ou non comme recommandation, essentiellement pour le cas de certaines phrases prises en contexte dans un Guide de Bonnes Pratiques. Au-delà des divergences, on pourra noter le rôle attribué à certaines constructions grammaticales comme l'emploi du futur. C'est ainsi le problème de la normalisation de la rédaction qui se profile. L'emploi du futur peut faire partie dans le sociolecte médical, des formes déontiques. Néanmoins, en l'absence d'étude approfondie de ce phénomène, pour rechercher en particulier l'influence du contexte, il semble prématuré d'attribuer systématiquement une interprétation déontique à la présence d'un verbe dénotant une action thérapeutique au futur ; le nombre de FP pourrait s'en trouver si augmenté que le système deviendrait inutilisable.

La section suivante s'intéresse plus particulièrement aux FP qui correspondent à des phrases reconnues par G-DEE comme des recommandations mais qui ne sont pas considérées comme telles par les experts.

9.2.7 Les FP détectées par des experts

Deux experts sur quatre pensent que l'Extrait 11 n'est pas une recommandation. L'un le justifie comme « Il n'y a pas d'étude. Ce n'est pas une recommandation, c'est une affirmation. C'est un constat. Ça ne veut pas dire qu'il faille définir un seuil. », alors que le dernier expert du panel pense que « c'est la justification de la recommandation qui précède. C'est plus un argumentaire. ». Mais, les deux autres experts considèrent cette phrase comme une recommandation. On touche encore une fois à un problème qui relève peut-être davantage du cycle de rédaction des Guides de Bonnes Pratiques.

Extrait 11

« Il n'existe pas d'étude permettant de définir un seuil tensionnel précis en dessous duquel un traitement antihypertenseur devrait être arrêté. »

L'Extrait 12 a suscité le rejet de la part d'un expert : « ce n'est pas une recommandation. ». L'interprétation de ce type d'affirmation consiste à faire intervenir le profil de l'expert et ce qui « pour lui » pourrait effectivement être perçu comme une recommandation.

Extrait 12

« Une évaluation de ces produits dans les 3 heures après l'AVC devrait être réalisée. »

Deux experts sur quatre considèrent que l'Extrait 13 est une recommandation, alors que les deux autres experts expriment leur réserve « On ne recommande pas de les utiliser. C'est une constatation. » et « C'est une information, à rapprocher conceptuellement de la phrase ».

9.2. *Evaluation par des experts*

Extrait 13

« Pour la mesure au cabinet, les appareils à mercure sont considérés comme les appareils de référence, cependant ils sont amenés à disparaître progressivement. »

L'Extrait 14 est également considéré comme une recommandation par deux experts sur quatre et la justification donnée par un expert la rejetant en tant que recommandation critiquant son caractère trop factuel « Ce n'est pas une recommandation. Ce sont des données de la littérature. »

Extrait 14

« Chez les diabétiques de type 2 ayant d'autres facteurs de risque cardiovasculaire, les IEC à dose suffisante pourraient apporter une protection cardiovasculaire spécifique (Grade B). »

Enfin, un expert sur quatre considère que l'Extrait 15 n'est pas une recommandation en le justifiant par « c'est un constat ».

Extrait 15

« La MAPA est plus particulièrement indiquée pour évaluer le mode de variation de la PA nocturne, pour rechercher une variabilité inhabituelle par exemple au cours d'une dysautonomie diabétique ou primitive, chez les patients atteints d'un syndrome d'apnée du sommeil ou chez les patients insuffisants rénaux chroniques et dans les cas où l'automesure n'est pas réalisable. »

Que ce soit pour les phrases FP ou FN, on observe une variabilité d'interprétation sur la notion de recommandation. Suite à une mise en commun de ces résultats, deux points de vue sont exposés : l'un voulant prendre position sur les phrases en désaccords permettant ainsi de catégoriser la phrase comme étant ou non une recommandation, l'autre consiste à ne pas chercher à résoudre les désaccords et donc à interpréter la raison pour laquelle une phrase est ou non considérée comme une recommandation.

Il semble que l'origine de certains désaccords se situe bien au-delà du contexte de ces expériences dans le processus de rédaction des Guides de Bonnes Pratiques. Ceci ouvre des perspectives intéressantes quant à la normalisation de l'écriture des Guides de Bonnes Pratiques. Plutôt que de proposer des normes de rédaction très strictes qui se heurteraient probablement à la réticence des rédacteurs, une étude ciblée des désaccords permettrait de proposer des règles spécifiques de rédaction destinées à éviter les cas d'ambiguïtés les plus flagrants.

9.3 Conclusion

A l'issue de ces expériences, il est nécessaire de dissocier les « erreurs » provenant effectivement de G-DEE, plus spécifiquement un problème de balisage des opérateurs déontiques pouvant concerner un problème de couverture syntaxique, des problèmes d'interprétation et de « qualité » d'écriture des recommandations.

L'évaluation a mis en évidence certains problèmes de structuration des Guides de Bonnes Pratiques. En effet, les experts avaient un point de vue différent sur les recommandations, bien que celles-ci aient déjà fait l'objet d'un consensus à l'intérieur du groupe de travail à l'origine de leur rédaction. Différentes notions sont ressorties de l'évaluation pour caractériser les phrases des Guides de Bonnes Pratiques. Par exemple, elles ont été caractérisées comme « une recommandation », « une information », « un constat », « une argumentation négative ». Les problèmes liés à la qualité d'écriture des recommandations entraînent une variabilité d'interprétation de ce qui constitue ou non une recommandation, pouvant être également expliquée par l'expérience et le point de vue des experts en identifiant une recommandation en fonction de leur vision de prescription.

L'évaluation a également permis de mettre en lumière quelques erreurs spécifiques de G-DEE (e.g. des limites de couverture syntaxique) qui concernent une recommandation sur onze recommandations problématiques¹⁰⁸ (pour le Guide de Bonnes Pratiques de l'AVC) et trois recommandations sur trente et une de problématiques (pour le Guide de Bonnes Pratiques de l'HTA). Une simple extension de la couverture syntaxique pourrait permettre une amélioration du taux de reconnaissance des opérateurs déontiques sans pour autant diminuer la précision.

Pour conclure ce chapitre, G-DEE pourrait être utilisé à l'avenir et sous une version étendue, pour l'aide à la rédaction des Guides de Bonnes Pratiques, et comme logiciel de support au consensus lors de leur écriture/rédaction. En effet, la mise en évidence de conflits lors de l'évaluation par les experts, permet de spécifier des outils d'aide au consensus lors de l'écriture des Guides de Bonnes Pratiques. Un autre axe issu de cette évaluation consiste à caractériser les problèmes provenant du passage « consensus oral » du groupe de travail à l'écriture textuelle des recommandations, ce qui permettrait d'améliorer potentiellement le document, tant au niveau de la structure que du contenu.

¹⁰⁸ On nomme ici les recommandations problématiques l'ensemble des recommandations représentées par FP et FN.

Chapitre 10

Conclusion

Ce chapitre est dédié aux perspectives d'utilisation de G-DEE après certaines améliorations et évolutions techniques. Ces perspectives sont présentées à travers le prisme de différents « modes d'utilisation » regroupant les fonctionnalités élémentaires en tâches de plus haut niveau intéressant également des populations d'utilisateurs spécifiques. Nous décrivons un ensemble de perspectives possibles à partir de G-DEE et proposant donc un G-DEE étendu pour intervenir à toutes les étapes du cycle de vie des Guides de Bonnes Pratiques, par exemple pour améliorer la rédaction des documents, ou faciliter leur informatisation.

10.1 Préambule

Nous nous sommes intéressés aux Guides de Bonnes Pratiques, qui sont des documents normalisés, nécessitant d'être standardisés et structurés pour faciliter leur informatisation mais aussi leur utilisation en pratique. Les approches documentaires basées sur les langages de balisage que nous avons détaillées dans le chapitre 2 se sont exclusivement attachées à la structuration du document comme préalable à son utilisation électronique. Nous avons proposé une extension de l'approche documentaire basée sur une analyse intelligente des marqueurs de structuration des recommandations utilisant un analyseur à états finis. Cette approche, intégrée à l'environnement informatique G-DEE (*Guideline Document Engineering Environment*), ouvre de multiples perspectives d'utilisation en automatisant certains aspects de la structuration, tout en restant compatible avec des standards *de facto* comme l'approche GEM (Shiffman et al., 2000).

Nous avons évalué les performances de G-DEE tant sur le plan technique (mesure de la qualité de la reconnaissance des marqueurs) que sur le plan pratique (évaluation par des experts de la structuration de recommandations élémentaires). Les performances techniques, qui correspondent essentiellement aux techniques empiriques de TALN sont très satisfaisantes pour ce type de problème avec environ 97% de succès. Les 3% restants ont été analysés plus en détail et ont montré qu'ils correspondaient le plus souvent à un défaut de couverture syntaxique de la forme des opérateurs déontiques. Les performances du système peuvent donc encore être améliorées, bien qu'il soit établi que ce type d'amélioration est de nature asymptotique. Dans l'ensemble, ces performances sont compatibles avec l'utilisation dans un système interactif où les erreurs de traitement résiduelles seront corrigées par l'utilisateur.

La seconde expérience que nous avons menée consistait à faire évaluer la structuration automatique de deux Guides de Bonnes Pratiques balisés par G-DEE par des experts (du domaine concerné mais aussi du domaine des Guides de Bonnes Pratiques). Les experts devaient valider les phrases balisées par G-DEE, i.e. le balisage de l'opérateur déontique en rouge, du *front-scope* en jaune et du *back-scope* en bleu, et vérifier que les phrases balisées correspondaient bien à des recommandations. Chaque grille de cotations remplie par les experts a ensuite été complétée par un entretien détaillé pour justifier l'interprétation de certaines phrases. Cette étape était rendue nécessaire par le fait que chaque expert avait évalué les phrases des Guides de Bonnes Pratiques en contexte, ce qui n'est pas sans conséquences quant aux jugements qu'ils ont formulés. Néanmoins, on a pu constater que le pourcentage de phrases correctement balisées par G-DEE et qui ont été validées par les experts varie entre 81 et 99% pour le Guide de Bonnes Pratiques de l'Accident Vasculaire Cérébral (AVC), et entre 83 à 91% pour celui de l'hypertension artérielle (HTA). Les résultats issus de cette évaluation préliminaire confirment la validité de l'approche et permettent de poser différentes problématiques quant à l'écriture et l'élaboration de ces documents textuels.

10.2 Impact potentiel de G-DEE sur le cycle de vie des Guides de Bonnes Pratiques

Nous présentons dans cette section, la place que G-DEE pourrait occuper¹⁰⁹, selon ses modes d'utilisation, dans le cycle de vie des Guides de Bonnes Pratiques, et les changements de pratiques qu'il pourrait entraîner.

Le premier mode d'utilisation de G-DEE concerne avant tout les professionnels en charge de l'élaboration des Guides de Bonnes Pratiques. Il consisterait à utiliser G-DEE comme un environnement d'étude pendant la rédaction des Guides de Bonnes Pratiques.

G-DEE supporte plusieurs affichages pour représenter les recommandations structurées à différents niveaux de détail, par exemple, l'affichage du *front-scope* et/ou du *back-scope*. On peut par exemple distinguer la structuration complète, i.e. l'affichage de l'opérateur déontique, du *front-scope* et du *back-scope*, qui permet d'analyser le texte dans son ensemble, par exemple pour des analyses de distribution des recommandations dans le texte, et pour vérifier la cohérence du document. Mais l'environnement graphique peut aussi permettre une étude plus fine de certaines parties des recommandations. Par exemple, l'affichage de l'opérateur déontique et du *front-scope*, ou de l'opérateur déontique et du *back-scope*. Ce type d'affichage permet de parcourir rapidement les textes à la recherche des actions diagnostiques ou thérapeutiques (auquel cas le maintien de l'opérateur déontique dans le texte surligné facilite la lisibilité).

Un deuxième mode d'utilisation de G-DEE concerne autant les experts en charge de l'élaboration des Guides de Bonnes Pratiques que ceux destinés à les relire (comme les membres du groupe de lecture). Ce mode nécessite la structuration complète des recommandations (décrite précédemment) et permet d'assister la lecture des Guides de Bonnes Pratiques par la mise en lumière des recommandations dans le texte, de la même manière qu'un document « surligné ». Le balisage automatique des recommandations permet donc une meilleure lisibilité et une identification plus rapide des recommandations pour discussion. Utilisé interactivement, il pourrait permettre d'apprécier en temps réel des variantes proposées du document.

Un autre mode d'utilisation de G-DEE consiste à assister la rédaction des Guides de Bonnes Pratiques et concerne plus particulièrement les experts en charge de leur élaboration. Ce mode se base sur l'analyse interactive du document et peut être utilisé pour l'homogénéisation des styles rédactionnels et la recherche d'ambiguïtés. Par exemple, certaines assertions reconnues comme des recommandations par des experts mais n'incorporant pas d'opérateurs déontiques lors de leur formulation initiale peuvent être modifiées pour faire apparaître ces opérateurs déontiques plus explicitement, non pas pour

¹⁰⁹ G-DEE reste un prototype de recherche et nécessiterait certains développements supplémentaires pour fonctionner en routine, par exemple la gestion d'erreurs générées par les utilisateurs, et l'optimisation des algorithmes implémentés pour améliorer les temps de réponses sur des documents complets.

la seule satisfaction de notre approche mais de façon plus pratique pour améliorer la lisibilité de la part de tous les publics des Guides de Bonnes Pratiques. Le texte formalisé GEM ou sous forme de règles de décision permet également de valider l'information écrite.

Un mode d'utilisation de G-DEE consiste à assister l'encodage des Guides de Bonnes Pratiques et intéresserait plus particulièrement les personnes en charge de leur informatisation. Ce mode nécessite une structuration plus fine des recommandations qui correspond par exemple au niveau de structuration décrit dans le formalisme GEM. Une phrase peut ainsi être analysée pour identifier les éléments correspondant aux variables de décision et aux actions du modèle GEM. À terme, G-DEE pourrait donc servir d'environnement semi-automatique pour l'encodage au standard GEM.

Le dernier mode d'utilisation que nous décrivons concerne l'acquisition des connaissances à partir de textes. Dans le cas des Guides de Bonnes Pratiques, les connaissances sous-jacentes peuvent être extraites pour alimenter des systèmes d'aide à la décision. G-DEE supporte la visualisation de règles de décision correspondant aux segments de texte sélectionnés et permet l'exploration et l'analyse du texte sous forme de règles SI – ALORS.

Nous pouvons conclure cette section par le fait que G-DEE, avec certaines améliorations, pourrait supporter en routine les différents stades du cycle de vie des Guides de Bonnes Pratiques. Nous présentons dans les sections suivantes des axes possibles suite aux travaux effectués autour de G-DEE, et selon les acteurs concernés.

10.3 Perspectives pour l'élaboration des Guides de Bonnes Pratiques

10.3.1 G-DEE pour l'aide à la rédaction

Dans le but d'apporter une aide au moment de l'écriture des Guides de Bonnes Pratiques, nous considérerons trois axes faisant intervenir : (i) la forme du document ; (ii) le contenu du document ; et (iii) la validation des recommandations écrites et du document.

Plusieurs études ont mis en évidence une variabilité dans la structuration des Guides de Bonnes Pratiques, tant au niveau de la forme que du contenu. Un premier axe de recherche consisterait à homogénéiser l'ensemble des documents produits par la même agence ou le même organisme sans pour autant imposer trop de contraintes aux rédacteurs.

Pour ce faire, deux approches sont possibles. La première (Figure 10.108) consiste à proposer interactivement des modèles de documents selon les sujets traités, comme le diagnostic, la thérapeutique et/ou le suivi d'une pathologie. Par exemple, en s'inspirant des modèles documentaires standard tels que GEM. Cette fonctionnalité pourrait être proposée aux chargés de projet en charge de structurer et d'écrire le contenu des Guides de Bonnes Pratiques.

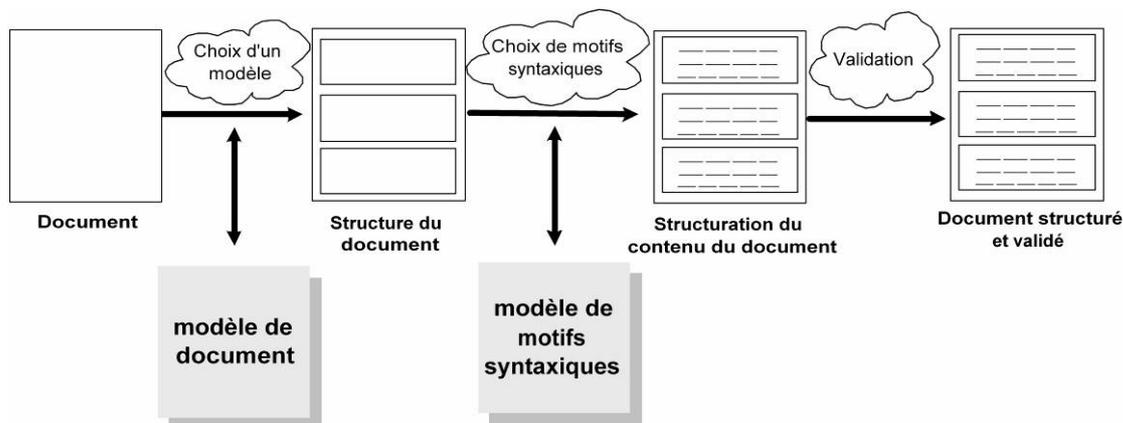


Figure 10.108 – Schéma représentant un exemple d'assistance à l'écriture des Guides de Bonnes Pratiques

La seconde approche consiste à identifier de manière automatique le « style » du texte, de vérifier la cohérence du style parmi l'ensemble des Guides de Bonnes Pratiques, puis de faire un retour à la personne en charge de la rédaction de ces documents des écarts qu'elle peut faire. Bien entendu l'analyse du style reposerait sur les mêmes hypothèses et les mêmes techniques que celles servant à la structuration, et le style serait défini en termes d'occurrence, de structure et de répartition des recommandations.

Une autre direction de recherche que nous n'avons pas abordée dans cette thèse, mais dans des travaux antérieurs, concerne l'adaptation du contenu des Guides de Bonnes Pratiques au cas des patients polypathologiques, souvent plus proches des « cas réels » que les praticiens rencontrent. Il faut souligner, qu'en règle générale, les Guides de Bonnes Pratiques sont développés suivant une approche « mono-pathologie » ce qui soulève un certain nombre de problèmes lorsqu'on recherche, par exemple, le traitement approprié d'un hypertendu diabétique atteint d'une insuffisance rénale. L'écriture spécifique de recommandations prenant en compte toutes les combinaisons possibles entre pathologies, même limitées aux tableaux les plus fréquents, serait une tâche trop complexe en pratique. C'est pourquoi, la possibilité de mettre en correspondance de façon automatique les recommandations de plusieurs Guides de Bonnes Pratiques « mono-pathologie » pourrait apporter un bénéfice réel. Cette consultation conjointe de plusieurs Guides de Bonnes Pratiques constituerait aussi une bonne illustration de l'utilité de leur informatisation.

Un deuxième axe de recherche concerne l'aide à l'écriture de recommandations structurées, en incluant la justification de ces recommandations correspondant, par exemple, aux bénéfices obtenus par le suivi d'une « bonne pratique ».

En ce qui concerne l'amélioration de la rédaction des recommandations elles-mêmes, on peut considérer deux approches. L'une peut s'inspirer de celle proposée par Guilbert (Guilbert, 1981) qui avait élaboré un guide décrivant les conduites à tenir pour la mise en œuvre des formations en Santé. Il proposait, en particulier, un ensemble de verbes actifs (*list of active verbs for stating educational objectives*) à utiliser dont voici un extrait :

Agir (<i>Act</i>)	Choisir (<i>Choose</i>)	Décider (<i>Decide</i>)
Administrer (<i>Administer</i>)	Décrire (<i>Describe</i>)	Diagnostiquer (<i>Diagnose</i>)
Appliquer (<i>Apply</i>)	Distinguer (<i>Distinguish</i>)	Faire (<i>Do</i>)
Assister (<i>Assist</i>)	Exécuter (<i>Execute</i>)	Guider (<i>Guide</i>)
Éviter (<i>Avoid</i>)	Justifier (<i>Justify</i>)	Modifier (<i>Modify</i>)
Changer (<i>Change</i>)	Obtenir (<i>Obtain</i>)	Réduire (<i>Reduce</i>)

On peut remarquer que certains verbes proposés correspondent aux verbes déontiques, tels que éviter et choisir. Dans le contexte actuel de normalisation des connaissances et d'homogénéisation du contenu des documents, nous pourrions penser à intégrer un module permettant une normalisation automatique des termes employés dans les Guides de Bonnes Pratiques. Néanmoins, ce type d'approche est difficilement réalisable, et a peu de chances d'être suivi en pratique. Ce qui nous amène à la seconde approche qui consiste à ne pas proposer de vocabulaire, mais à interagir directement avec l'utilisateur de G-DEE pour lui signaler qu'une phrase pourrait être mal comprise du fait de sa tournure ou de sa complexité. Par exemple, le cas d'une recommandation complexe de la forme SI ... ALORS ... SAUF :

Un traitement antiplaquettaire par aspirine (à la posologie de 160 à 300 mg/j) est recommandé dès que possible après un AVC ischémique (grade A), **sauf** si un traitement thrombolytique est envisagé.

La mise en évidence de conflits lors de l'évaluation par les experts a mis encore une fois en évidence l'importance d'un accord sur l'interprétation des recommandations. Le processus d'élaboration d'un Guide de Bonnes Pratiques repose en grande partie sur un consensus d'experts, pour lequel on peut noter, comme décrit dans le chapitre 9, que la recommandation constitue une interprétation directe en terme d'acte de langage même en l'absence d'opérateurs déontiques, par exemple, l'emploi du futur. Un axe de recherche prometteur consisterait donc à spécifier des outils d'aide au consensus lors de l'écriture des Guides de Bonnes Pratiques, intégré à l'environnement G-DEE.

D'autres axes de recherche pourraient s'orienter vers la qualité d'écriture des recommandations en fonction de grilles déjà existantes, telles que le modèle COGS (Shiffman & Michel, 2004), récemment proposé et approuvé comme un document standard pour énumérer les éléments relatifs à la qualité d'un Guide de Bonnes Pratiques. Ce modèle reprend un sous-ensemble des éléments proposés dans le modèle GEM, et récemment, eGLIA a été proposé comme une interface informatique, sur la base du modèle COGS permettant de valider en temps réel, par plusieurs experts, la qualité des Guides de Bonnes Pratiques (Shiffman et al., 2005). La synthèse des accords/désaccords est ensuite générée automatiquement. Dans le même sens, AGREE¹¹⁰ (*Appraisal of Guidelines Research & Evaluation*) (AGREE, 2003) est aussi une grille servant à évaluer la qualité d'un Guide de Bonnes Pratiques, plus précisément à identifier les biais potentiels lors du développement du document. La grille contient des critères spécifiques, tels que la

¹¹⁰ <http://www.agreecollaboration.org/> (dernier accès le 10 juin 2006).

portée et le but du Guide de Bonnes Pratiques, la rigueur de développement, la clarté et la présentation du document. Ces grilles sont instanciées manuellement et leur remplissage automatique pourrait faciliter leur utilisation et leur adoption.

Dans le cas des modèles proposés, on évalue le Guide dans son entier, alors que nous cherchons plus précisément à évaluer les recommandations élémentaires elles-mêmes une à une. Dans ce contexte, nous pourrions proposer des méthodes de validation des recommandations par un indicateur de qualité, par exemple, en considérant des mesures comme le ratio des phrases balisées par G-DEE par rapport au texte entier. On pourrait imaginer qu'un Guide ne contenant qu'un faible pourcentage ne représenterait pas forcément un Guide applicable en pratique, ou manquant cruellement de recommandations, ou pouvant être très peu spécifique, ou ne contenant que très peu de conduites à tenir. À l'inverse, ce faible pourcentage pourrait représenter le fait que le Guide contient énormément de « bruit » et que les recommandations sont les éléments pouvant être intégrés dans une fiche de synthèse. Ainsi, un autre axe ciblerait spécifiquement l'élaboration de méthodes de génération automatique de fiches de synthèse à partir de Guides de Bonnes Pratiques structurés, en indiquant le pourcentage de couverture par rapport au document initial.

En conclusion, G-DEE pourrait être utilisé à l'avenir et sous une version étendue, pour l'aide à la rédaction des Guides de Bonnes Pratiques, et comme logiciel de support au consensus lors de leur écriture/rédaction.

10.3.2 G-DEE pour l'aide à la maintenance et à l'actualisation

Les Guides de Bonnes Pratiques sont régulièrement actualisés pour refléter l'évolution des connaissances, l'impact de nouvelles études ou même des modifications d'environnement socio-économique du système de soins. Il est donc potentiellement intéressant de proposer des outils qui puissent faciliter la mise-à-jour des Guides de Bonnes Pratiques, favorisant par exemple l'intégration de modifications tout en maintenant les propriétés des documents.

Un premier axe consisterait à proposer des méthodes semi-automatiques d'aide à l'actualisation des Guides de Bonnes Pratiques ; celles-ci pourraient utiliser des techniques d'acquisition des connaissances à partir des textes pour comparer le contenu des Guides de Bonnes Pratiques à un instant donné à celui d'autres bases de connaissances. A minima, on pourrait envisager de sélectionner automatiquement selon des critères définis les derniers essais cliniques pour pouvoir les comparer avec le contenu des Guides de Bonnes Pratiques et observer si cette connaissance amène ou non à de nouvelles conduites à tenir.

Un objectif plus ambitieux consisterait à générer directement des recommandations à partir des essais cliniques. Cette phase consisterait en une première structuration des recommandations, et améliorerait la structuration des recommandations en contexte. Elle pourrait cependant se heurter elle aussi à des problèmes de gestion de consensus et de fusion des informations provenant de différentes sources.

10.4 Perspectives pour l'informatisation des Guides de Bonnes Pratiques

Le caractère normatif et validé des Guides de Bonnes Pratiques a suscité un intérêt pour leur utilisation dans le domaine de l'acquisition des connaissances, en particulier pour les systèmes d'aide à la décision. L'informatisation du contenu des Guides de Bonnes Pratiques permet de les utiliser comme sources de connaissances mais on peut également envisager des systèmes d'aide à la décision destinés principalement à la promotion des bonnes pratiques. L'extraction de connaissances prend comme point de départ les recommandations pour produire des représentations informatiques comme les règles de décision. Dans cette optique, une fonctionnalité d'un « G-DEE étendu » consisterait à s'intéresser au contenu textuel des *scopes* pour identifier automatiquement, par exemple, les traitements médicaux par la reconnaissance de descriptions ou *Named Entity Recognition* (NER). Cette fonctionnalité faciliterait l'étape d'informatisation des règles de décision, par exemple pour caractériser et normaliser les connaissances avant leur intégration dans des systèmes d'aide à la décision.

Dans un contexte similaire, le projet EsPeR¹¹¹ « Estimation Personnalisée des Risques en santé » (Colombet et al., 2005) développé au sein de notre laboratoire est un système d'aide à la décision spécifique à la prévention. EsPeR est accessible à travers un site Internet destiné aux médecins libéraux et propose une estimation individualisée, quantifiée et reproductible des risques de mortalité par cause, ainsi qu'une estimation du risque chiffrée ou semi quantitative dans le domaine des maladies cardiovasculaires (risque cardiovasculaire global), de certains cancers, du suicide et de la consommation d'alcool.

À titre d'illustration, les arbres de décision d'EsPeR basés sur le modèle GLIF (formalisme décrit dans le chapitre 2) pourraient être obtenus à partir du balisage de G-DEE. Cette étape intégrée au projet a permis d'identifier plus rapidement les recommandations et de faciliter la construction de l'arbre de décision de manière reproductible pour l'ensemble des Guides de Bonnes Pratiques d'EsPeR. Des travaux sont actuellement en cours pour générer automatiquement des arbres de décision en fonction des règles de décision de G-DEE. Un axe de recherche consiste à caractériser les règles potentiellement pertinentes et nécessaires pour la construction d'arbres de décision. Cette approche intervenant aux différentes étapes du cycle de vie dans EsPeR a l'ambition de rendre le processus d'élaboration d'arbres de décision plus automatique et surtout plus reproductible.

¹¹¹ <http://www.hegp.bhdc.jussieu.fr/esper/> (dernier accès le 10 juin 2006).

Des travaux futurs dans ce sens pourraient s'orienter vers la construction d'arbres basés sur des techniques de compréhension profonde du texte.

10.5 Conclusion générale

Nous nous sommes plus particulièrement attachés au Guide de Bonnes Pratiques en tant que document et à la manière dont nous pourrions l'améliorer en proposant des outils d'aide à la rédaction, au consensus, à l'informatisation.

Au cours de nos travaux, nous avons pu constater que les Guides de Bonnes Pratiques ont besoin d'être structurés pour être mieux utilisés et mieux acceptés par les médecins. Cette structuration reste aussi nécessaire lors de leur informatisation, et constitue une étape commune à la conception d'un Guide de Bonnes Pratiques « papier » ou « numérique ».

L'approche que nous avons adoptée consiste à proposer une méthode d'analyse semi-automatique des Guides de Bonnes Pratiques, basée sur la notion d'opérateurs déontiques, pouvant agir directement sur les problèmes de structuration, de cohérence et d'interprétation. Des outils d'assistance informatique peuvent donc :

- aider à révéler les incohérences et/ou incomplétudes du document ;
- aider à améliorer la clarté du document, par exemple les critères de décision et les recommandations cliniques ;
- proposer un support au consensus ;
- proposer un support au cycle de vie du document intégrant les fonctionnalités ci-dessus.

Tous ces points, non encore résolus actuellement, nous laissent penser que les recherches sur les Guides de Bonnes Pratiques sont appelées à se développer, tant en raison de l'importance que ces documents sont susceptibles d'acquérir dans un avenir proche qu'en raison des multiples besoins qu'ils suscitent et qui peuvent être pourvus par des approches informatiques bien conçues.

Références

AGRAWAL, A. & MAYO-SMITH, M. (2004). Adherence to Computerized Clinical Reminders in a Large Healthcare Delivery Network. In *IMLA* (ed. M. e. a. Fieschi), pp. 111-114. Amsterdam: IOS Press, San Francisco.

AGREE, C. (2003). Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care*, **12**, 18-23.

AKHTAR, S., REILLY, R. & DUNNION, J. (2003). Automating XML markup of text documents. In *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers*, vol. 2, pp. 1-3.

ANAES. (2000). *Efficacité des méthodes de mise en oeuvre des recommandations médicales*, Paris.

ANAND, V., BIONDICH, P., LIU, G., ROSENMAN, M. & DOWNS, S. (2004). Child Health Improvement through Computer Automation: The CHICA System. In *IMLA* (ed. M. F. e. al.), pp. 187-191. Amsterdam: IOS Press, San Francisco.

ARMSTRONG, D., TATFORD, P., FRY, J. & ARMSTRONG, P. (1992). Development of clinical guidelines in a health district: an attempt to find consensus. *Qual Health Care* **1**, 241-244.

AUSTIN, J. L. (1962). *How to do things with words*, 1970 edition. Le Seuil, Paris.

BANG, M. & ERIKSSON, H. (1997). Generation of development environments for the Arden Syntax. In *Proceedings of the AMLA Symposium*, pp. 313-317.

BASIL, R., GRISOLI, F. & PAZIENZA, M. (1994). Might a semantic lexicon support hypertextual authoring? In *Proceedings of the fourth conference on Applied Natural Language Processing*, pp. 174-179.

BENNETT, J. & GLASZIOU, P. (2003). Computerised reminders and feedback in medication management: a systematic review of randomised controlled trials. *MJA* **178**, 217-222.

BERNSTAM, E., ASH, N., PELEG, M., TU, S., BOXWALA, A., MORK, P., SHORTLIFFE, E. & GREENES, R. (2000). Guideline classification to assist modeling, authoring, implementation and retrieval. In *Proc AMLA Symp. 2000*, pp. 66-70.

BERO, L., GRILLI, R., GRIMSHAW, J., HARVEY, E., OXMAN, A. & THOMSON, M. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ* **317**, 465-8.

BINDELS, R., DE CLERCQ, P., WINKENS, R. & HASMAN, A. (2000). A test ordering system with automated reminders for primary care based on practice guidelines. *International Journal of Medical Informatics* **58-59**, 219-233.

BOXWALA, A., TU, S., PELEG, M., ZENG, Q., OGUNYEMI, O., GREENES, R., SHORTLIFFE, E. & PATEL, V. (2001). Towards a representation format for sharable clinical guidelines. *J Biomed Inform* **34**, 157-169.

BURSTIN, H., CONN, A., SETNIK, G., RUCKER, D., CLEARY, P., O'NEIL, A., ORAV, E., SOX, C. & BRENNAN, T. (1999). Benchmarking and quality improvement: the Harvard Emergency Department Quality Study. *Am J Med* **107**, 437-49.

CABANA, M., RAND, C., POWE, N., WU, A., WILSON, M., ABBOUD, P. & RUBIN, H. (1999). Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement. *JAMA* **282**, 1458-1465.

CAMPBELL, J., TU, S., MANSFIELD, J., BOYER, J., MCCLAY, J., PARKER, C., RAM, P., SCHEITEL, S. & McDONALD, K. (2003). The SAGE Guideline Model: A Knowledge Representation Framework for Encoding Interoperable Clinical Practice Guidelines. Stanford Medical Informatics Report SMI-2003-096. http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-2002-0962.html

CANIARD, E. (2002). Les Recommandations de bonnes pratiques : un outil de dialogue, de responsabilité et de diffusion de l'information. Rapport à M. Bernard Kouchner.

CICCARESE, P., CAFFI, E., BOIOCCHI, L., HALEVY, A., QUAGLINI, S., KUMAR, A. & STEFANELLI, M. (2003). The NewGuide Project: guidelines, information sharing and learning from exceptions. In *Artificial Intelligence in Medicine*, pp. 163-167. Springer, Cyprus.

CICCARESE, P., CAFFI, E., BOIOCCHI, L., QUAGLINI, S. & STEFANELLI, M. (2004). A Guideline Management System. In *IMLA* (ed. M. F. e. al.), pp. 28-32. Amsterdam: IOS Press, San Francisco.

CICCARESE, P., CAFFI, E., QUAGLINI, S. & STEFANELLI, M. (2005). Architectures and tools for innovative Health Information Systems: The Guide Project. *Int J Med Inform* **74**, 553-62.

CODISH, S. & SHIFFMAN, R. (2005). A Model of Ambiguity and Vagueness in Clinical Practice Guideline Recommendations. In *Proc AMLA 2005* (ed. C. Friedman), pp. 146-150, Washington DC.

COLOMBET, I., AGUIRRE-JUNCO, A., ZUNINO, S., JAULENT, M., LENEVEUT, L. & CHATELLIER, G. (2005). Electronic implementation of guidelines in the EsPeR system: a knowledge specification method. *Int J Med Inform* **74**, 597-604.

COSTANZA, M., STODDARD, A., ZAPKA, J., GAW, V. & BARTH, R. (1992). Physician compliance with mammography guidelines: barriers and enhancers. *J Am Board Fam Pract* **5**, 143-52.

COWIE, J., GUTHRIE, L. & WILKS, Y. (1992). CRL/NMSU and Brandeis: Description of the *MucBruce* System as Used for MUC-4. In *MUC-4: Fourth Message Understanding Conference* (ed. D. A. R. P. Agency and S. a. I. S. T. Office), pp. 223-232. Morgan Kaufmann, McLean, Virginia.

COWIE, J. & WILKS, Y. (2000). *Information Extraction, A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*, Editors: Robert Dale, Hermann Moisl, and Harold Somers, Marcel Dekker Inc.

CRANNEY, M., WARREN, E., BARTON, S., GARDNER, K. & WALLEY, T. (2001). Why do GPs not implement evidence-based guidelines? A descriptive study. *Family Practice* **18**, 359-363.

DARPA SOFTWARE AND INTELLIGENT SYSTEMS TECHNOLOGY OCE. (1992). Proceedings of the Fourth Message Understanding Conference (MUC-4), McLean, Virginia, June 1992. DARPA Software and Intelligent Systems Technology Office, Morgan Kaufmann, San Mateo, California.

DAWES, M. (1996). On the need for evidence-based general and family practice. *Evidence-based Medicine* **1**, 68-69.

DE CLERCQ, P. (2000). An ontological approach for the development of shareable guidelines. In *Proceedings of AMLA Symposium*, pp. 166:170.

DE CLERCQ, P. (2003). Guideline-based Decision Support in Medicine Modeling Guidelines for the Development and Application of Clinical Decision Support Systems, Technische Universiteit Eindhoven.

DE CLERCQ, P., HASMAN, A., BLOM, J. & KORSTEN, H. (2001). Design and implementation of a framework to support the development of clinical guidelines. *Int. J. Med. Inform.* **64**, 285-318.

DRILLON, J. (1991). *Traité de la ponctuation française*, Gallimard edition, Paris.

DYMETMAN, M., LUX, V. & RANTA, A. (2000). XML and multilingual document authoring: convergent trends. In *Proceedings of the 18th conference on Computational linguistics*, vol. 1, pp. 243-249. Association for Computational Linguistics.

ELKIN, P., PELEG, M., LACSON, R., BERNSTAM, E., TU, S., BOXWALA, A., GREENES, R. & SHORTLIFFE, E. (2000). Toward standardization of electronic guidelines. *MD Comput.* **17**, 39-44.

ERIKSSON, H. & MUSEN, M. (1993). Metatools for knowledge acquisition. *IEEE Software* **10**, 23-29.

FARQUHAR, C., KOFA, E. & SLUTSKY, J. (2002). Clinicians' attitudes to clinical practice guidelines: a systematic review. *MJA* **177**, 502-506.

FARQUHAR, C., KOFA, E. & SLUTSKY, J. (2003). Clinicians report high satisfaction with clinical guidelines, but question their practicability. *Evidence-based Healthcare* **7**, 77-78.

FERRES, D., KANAAN, S., DOMINGUEZ-SAL, D., GONZALEZ, E., AGENO, A., FUENTES, M., RODRIGUEZ, H., SURDEANU, M. & TURMO, J. (2005). TALP-UPC at TREC 2005: Experiments Using a Voting Scheme Among Three Heterogeneous QA Systems. In *The Fourteenth Text REtrieval Conference (TREC)*. http://trec.nist.gov/pubs/trec14/t14_proceedings.html

FIELD, M. & LOHR, K. (1990). *Guidelines for clinical practice: Directions for a new program*. National Academy Press, Washington DC.

FOX, J., BEVERIDGE, M. & GLASSPOOL, D. (2003). Understanding intelligent agents: analysis and synthesis. *Aicommunications* **16**, 139-152(14).

FOX, J. & RAHMANZADEH, A. (1998). Disseminating medical knowledge: the PROforma approach. *Artificial Intelligence in Medicine* **14**, 157-181.

FRANKEL, H., FITZPATRICK, M., GASKELL, S., HOFF, W., ROTONDO, M. & SCHWAB, C. (1999). Strategies to improve compliance with evidence-based clinical management guidelines. *J Am Coll Surg* **189**, 533-538.

GARDNER, R., PRYOR, T. & WARNER, H. (1999). The HELP Hospital Information system: up-date 1998. *International Journal of Medical Informatics* **54**, 169-182.

GARG, A., ADHIKARI, N., MCDONALD, H., ROSAS-ARELLANO, M., DEVEREAUX, P., BEYENE, J., SAM, J. & HAYNES, R. (2005). Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes - A Systematic Review. *JAMA* **293**, 1223-1238.

GEORG, G. (2005). Computerization of Clinical Guidelines: an Application of Medical Document Processing. In *Intelligent Paradigms for Healthcare Enterprises Systems Thinking*, vol. 184. *Studies in Fuzziness and Soft Computing* (ed. B. G. J. Silverman, A.; Ichalkaranje, A.; Jain, L.C.), pp. 1-30. Springer-Verlag.

GEORG, G., COLOMBET, I. & JAULENT, M.-C. (2005). Structuring Clinical Guidelines through the Recognition of Deontic Operators. In *Proceedings of Medical Informatics Europe 2005*, vol. 116, pp. 151-156. IOS Press Amsterdam, Geneva, Switzerland.

GEORG, G. & JAULENT, M.-C. (2005). An Environment for Document Engineering of Clinical Guidelines. In *Proceedings of the American Medical Informatics Association*, pp. 276-280, Washington, DC.

GEORG, G., SÉROUSSI, B. & BOUAUD, J. (2003). Dérivation d'une base de connaissances à partir d'une instance GEM d'un guide de bonnes pratiques médicales textuel. In *14e journées francophones d'Ingénierie des Connaissances* (ed. D.-K. R), pp. 51-66. PUG, Laval.

GEORG, G., SÉROUSSI, B. & BOUAUD, J. (2005). Extending the GEM model to support knowledge extraction from textual guidelines. *International Journal of Medical Informatics* **74**, 79-87.

GERSHKOVICH, P. & SHIFFMAN, R. (2001). An implementation framework for GEM encoded guidelines. In *Proc AMLA Symp* (ed. B. S), pp. 204-208.

GIARRATANO, J. & RILEY, G. (1994). *Expert Systems: Principles and Programming*. PSW Publishing Company, Boston.

GILLOIS, P., CHATELLIER, G., JAULENT, M., COLOMBET, I., FIESCHI, M. & DEGOULET, P. (2001). From paper-based to electronic guidelines : application to French guidelines. In *Medinfo*, vol. 10(Pt 1), pp. 196-200.

GOLDFARB, C. (1981). A generalized approach to document markup. In *ACM SIGPLAN SIGOA symposium on Text manipulation*, vol. 16, pp. 68-73. ACM Press.

GREENES, R., BOXWALA, A., SLOAN, W., OHNO-MACHADO, L. & SR, D. (1999). A framework and tools for authoring, editing, documenting, sharing, searching, navigating, and executing computer-based clinical guidelines. In *Proc AMLA Symp*, vol. 1-2, pp. 261-265.

GREENES, R., PELEG, M., BOXWALA, A., TU, S., PATEL, V. & SHORTLIFFE, E. (2001). Sharable Computer-Based Clinical Practice Guidelines: Rationale, Obstacles, Approaches, and Prospects. In *Medinfo*, vol. 10, pp. 201-205.

- GRENIER, B. (1990, 1996). *Evaluation de la décision médicale - Introduction à l'analyse médico-économique*, 2e édition edition. Masson, Paris.
- GRIMSHAW, J. & ECCLES, M. (2004). Is evidence-based implementation of evidence-based care possible? *MJA* **180**, S50-S51.
- GRIMSHAW, J. & RUSSEL, I. (1993). Effects of clinical guidelines on medical practice: a systematic review of rigorous evaluation. *Lancet* **342**, 1317–1322.
- GUILBERT, J. (1981). *Educational Handbook for Health Personnel*. World Health Organization, Geneva.
- HAGERTY, C., CHANG, J., PICKENS, D., KULIKOWSKI, C. & SONNENBERG, F. (2004). Semi-automated Encoding of Guidelines. In *Medinfo*, pp. 1625, San Francisco.
- HAGERTY, C., PICKENS, D., KULIKOWSKI, C. & SONNENBERG, F. (2000). HGML: A Hypertext Guideline Markup Language. In *Proc. Annual Meeting of the American Medical Informatics Association (AMIA)*, pp.325-329, Los Angeles.
- HERBERT, S., GORDON, C., JACKSON-SMALE, A. & SALIS, J. (1995). Protocols for clinical care. *Comput Methods Programs Biomed* **48**, 21–26.
- HEWITT-TAYLOR, J. (2003). Reviewing evidence. *Intensive and Critical Care Nursing* **19**, 43-49.
- HEWITT-TAYLOR, J. (2004). Clinical guidelines and care protocols. *Intensive and Critical Care Nursing* **20**, 45-52.
- HIRSCHMAN, L. & CHINCHOR, N. (1998). Muc-7 named entity task definition. In *Proceedings of the 7th Message Understanding Conference*.
- HOBBS, J., APPELT, D. & BEAR, J. (1991). Robust Processing of Real-World Natural-Language Texts. In *SRI International*, pp. 1-24.
- HOBBS, J., APPELT, D., BEAR, J., ISRAEL, D., KAMEYAMA, M., STICKEL, M. & TYSON, M. (1996). FASTUS: A cascaded finite-state transducer for extracting information from natural language text. In *Finite State Devices for Natural Language Processing* (ed. I. R. a. Schabes.), pp. 383-406. MIT Press.
- HRIPCSAK, G. & ROTHSCHILD, A. (2005). Agreement, the F-Measure, and Reliability in Information Retrieval. *J Am Med Inform Assoc* **12**, 296–298.

JENDERS, R., CORMAN, R. & SASGUPTA, B. (2003). Making the standard more standard: a data and query model for knowledge representation in the Arden Syntax. In *Proceedings of the Amer Med Informatics Assoc*, pp. 323- 327. Musen M., Washington, DC.

JENDERS, R. & DASGUPTA, B. (2002). Challenges in implementing a knowledge editor for the Arden Syntax: knowledge base maintenance and standardization of database linkages. In *Proc AMLA Symp.*, pp. 355-359.

JOHNSON, P., TU, S., BOOTH, N., SUGDEN, B. & PURVES, I. (1999). A Guideline Model for Chronic Disease Management in Primary Care. Report SMI-1999-790, Medical Informatics, Stanford University School of Medicine, 1999.

JOHNSON, P., TU, S., BOOTH, N., SUGDEN, B. & PURVES, I. (2000). Using scenarios in chronic disease management guidelines for primary care. In *Proc AMLA Symp*, pp. 389-393.

JOUSIMAA, J., MAKELA, M., KUNNAMO, I. & ET, A. (2003). Computerised and paper guidelines may have similar effects on consultation practices among newly qualified physicians. *Evidence-based Healthcare* **7**, 79-80.

KALINOWSKI, G. (1996). *La Logique Déductive*. Presses Universitaires de France (in French).

KARRAS, B., NATH, S. & SHIFFMAN, R. (2000). A Preliminary Evaluation of Guideline Content Mark-up Using GEM--An XML Guideline Elements Model. In *Proc AMLA Symp*, pp. 413-417.

KAVANAGH, M. (2002). The quest for a computerized guideline standard: The process, its history, and an evaluation of the most common and promising methods used today.

KECKLEY, P. (2004). Evidence-Based Medicine in Managed Care: a Survey of Current and Emerging Strategies. *Medscape General Medicine* **6**.

KEFFER, J. (2001). Guidelines and Algorithms: Perceptions of Why and When They Are Successful and How to Improve Them. *American Association for Clinical Chemistry* **47**, 1563-1572.

KOSARA, R. & MIKSCH, S. (2001). Metaphors of movement: a visualization and user interface for time-oriented, skeletal plans. *Artif. Intell. Med.* **22**, 111-131.

KOSARA, R., MIKSCH, S., SEYFANG, A. & VOTRUBA, P. (2002). Tools for Acquiring Clinical Guidelines in Asbru. In *Integrated Design and Process Technology*, pp. 1-6.

- KUMAR, A., SMITH, B., PISANELLI, D., GANGEMI, A. & STEFANELLI, M. (2004). An ontological framework for the implementation of clinical guidelines in health care organizations. In *Stud Health Technol Inform*, vol. 102, pp. 95-107.
- KUPERMAN, G., GARDNER, R. & PRYOR, T. (1991). *The HELP System*. Springer-Verlag, New-York.
- LEE, G., SEO, J., LEE, S., JUNG, S., CHO, B.-H., LEE, C., KWAK, B.-K., CHA, J., KIM, D., AN, J., KIM, H. & KIM, K. (2001). SiteQ: Engineering High Performance QA system Using Lexico-Semantic Pattern Matching and Shallow NLP. In *The 10 Text Retrieval Conference (TREC-10)*, pp. 437-446, Maryland.
- LOBACH, D. & HAMMOND, W. (1997). Computerized decision support based on a clinical practice guideline improves compliance with care standards. In *Am J Med*, vol. 102, pp. 89-98.
- LOMAS, J. (1991). Words without action? The production, dissemination, and impact of consensus recommendations. *Annu Rev Public Health* **12**, 41-65.
- MATILLON, Y. & DURIEUX, P. (2000). L'évaluation médicale, du concept à la pratique (ed. F. Médecine-Sciences), pp. 43-54.
- MICALISTER, F., GRAHAM, I., KARR, G. & LAUPACIS, A. (1999). Evidence-Based Medicine and the Practicing Clinician. *J Gen Intern Med* **14**, 236-242.
- MCGLYNN, E., ASCH, S., ADAMS, J., KEESEY, J., HICKS, J., DECRISTOFARO, A. & KERR, E. (2003). The Quality of Health Care Delivered to Adults in the United States. *The New England Journal of Medicine* **348**, 2635-2645.
- MIDWEST BUSINESS GROUP ON HEALTH. (2003). Reducing the Costs of Poor-Quality Health Care Through Responsible Purchasing Leadership (ed. I. M. B. G. o. H. Chicago).
- MIKSCH, S., KOSARA, R., SHAHAR, Y. & JOHNSON, P. (1998). AsbruView: visualization of time-oriented, skeletal plans. In *Proceedings of the Fourth International Conference on Artificial Intelligence Planning Systems.*, pp. 11–18, Pittsburgh, PA: Carnegie-Mellon University.
- MIKSCH, S., SHAHAR, Y. & JOHNSON, P. (1997). Asbru: a task-specific, intention-based, and time-oriented language for representing skeletal plans. In *Proceedings of the Seventh Workshop on Knowledge Engineering Methods and Languages (KEML-97)*. Milton Keynes, UK.
- MOLDOVAN, D., CHA, S., CHUNG, M., GALLIPPI, T., HENDRICKSON, K., KIM, J., LIN, C. & LIN, C. (1993). USC: Description of the SNAP system used for MUC-5. In *The 5th conference on Message understanding*, pp. 305-319.

MOSER, M. (2004). Identifying Pattern in Clinical Guidelines, Institut für Softwaretechnik der Technischen Universität Wien.

MOULIN, B. & ROUSSEAU, D. (1990). Knowledge acquisition from prescriptive texts. In *International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-90*, pp. 1112-1121, Charleston.

MUSEN, M., GENNARI, J., ERIKSSON, H., TU, S. & PUERTA, A. (1995). PROTEGE-II: computer support for development of intelligent systems from libraries of components. In *Medinfo*, vol. 8, pp. 766-770.

OHNO-MACHADO, L., GENNARI, J., MURPHY, S., JAIN, N., TU, S., OLIVER, D., PATTISON-GORDON, E., GREENES, R., SHORTLIFFE, E. & BARNETT, G. E. A. (1998 Jul - Aug). The guideline interchange format: a model for representing guidelines. *Journal of the American Medical Informatics Association* **5**, 357-372.

PATEL, V., ALLEN, V., AROCHA, J. & SHORTLIFFE, E. (1998 Sep-Oct). Representing clinical guidelines in GLIF: individual and collaborative expertise. *J Am Med Inform Assoc* **5**, 467-83.

PATEL, V., AROCHA, J., DIERMEIER, M., GREENES, R. & SHORTLIFFE, E. (2001a). Methods of cognitive analysis to support the design and evaluation of biomedical systems: the case of clinical practice guidelines. *J Biomed Inform* **34**, 52-66.

PATEL, V., AROCHA, J., DIERMEIER, M., HOW, J. & MOTTUR-PILSON, C. (2001b). Cognitive psychological studies of representation and use of clinical practice guidelines. *Int J Med Inf* **63**, 147-67.

PATEL, V., BRANCH, T., MOTTUR-PILSON, C. & PINARD, G. (2004). Public Awareness About Depression: The Effectiveness of a Patient Guideline. *International Journal of Psychiatry in Medicine* **34**, 1-20.

PAZIENZA, M. T. (1997). *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*. Springer-Verlag.

PELEG, M., BOXWALA, A., OGUNYEMI, O., ZENG, Q., TU, S., LACSON, R., BERNSTAM, E., ASH, N., MORK, P., OHNO-MACHADO, L., SHORTLIFFE, E. & GREENES, R. (2000). GLIF3: The Evolution of a Guideline Representation Format. In *Proc AMLA Symp*, pp. 645-649.

PELEG, M. & KANTOR, R. (2003). Approaches for Guideline Versioning Using GLIF. In *Proc AMLA Symp*, pp. 509-513.

PELEG, M., OGUNYEMI, O., TU, S., BOXWALA, A., ZENG, Q., GREENES, R. & SHORTLIFFE, E. (2001). Using features of Arden Syntax with object-oriented medical data models for guideline modeling. In *Proc AMLA Symp*, pp. 523-527.

- PELEG, M., TU, S., BURY, J., CICCARESE, P., FOX, J., GREENES, R., HALL, R., JOHNSON, P., JONES, N., KUMAR, A., MIKSCH, S., QUAGLINI, S., SEYFANG, A., SHORTLIFFE, E. & STEFANELLI, M. (2003). Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *J Am Med Inform Assoc* **10**, 52-68.
- PERRY, S. & MARX, E. (1994). Practice guidelines: role of economics and physicians. *Health Aff (Millwood)* **13**, 141-145.
- PISANELLI, D., GANGEMI, A. & STEVE, G. (1999). Towards a Standard for Guideline Representation: an Ontological Approach. *JAMIA* **6**, 906-910.
- PRYOR, T. & HRIPCSAK, G. (1993). The Arden Syntax for Medical Logic Modules. *International Journal of Clinical Monitoring and Computing* **10**, 215-224.
- PURVES, I., SUGDEN, B., BOOTH, N. & SOWERBY, M. (1999). The PRODIGY project—the iterative development of the release one model. In *Proceedings of the AMLA Symposium*, pp. 359-363.
- QUAGLINI, S., STEFANELLI, M., CAVALLINI, A., MICIELI, G., FASSINO, C. & MOSSA, C. (2000). Guideline-based careflow systems. *Artif. Intell. Med.* **20**, 5–22.
- QUAGLINI, S., STEFANELLI, M., LANZOLA, G., CAPORUSSO, V. & PANZARASA, S. (2001). Flexible guideline-based patient careflow systems. *Artificial Intelligence in Medicine* **22**, 65-80.
- RAM, P., BERG, D., TU, S., MANSFIELD, G., YE, Q., ABARBANEL, R. & BEARD, N. (2004). Executing Clinical Practice Guidelines using the SAGE Execution Engine. In *IMLA* (ed. M. F. e. al.), pp. 251-255. Amsterdam: IOS Press, San Francisco.
- RASTIER, F. (1989). *Sens et textualité*. Hachette, Paris.
- RASTIER, F. (2000). L'accès aux banques textuelles - Des genres à la doxa. In *Lèxic, corpus i diccionaris* (ed. C. Cabré T. et Gelpi), IULA, Universitat Pompeu Fabra, Barcelone.
- ROCHE, E. & SCHABES, Y. (1997). *Finite-State Language Processing*. MIT Press.
- SACKETT, D., ROSENBERG, W., GRAY, J., HAYNES, R. & RICHARDSON, W. (1996). Evidence-based medicine: what it is and what it isn't. *BMJ* **312**, 71-72.
- SEARLE, J. (1965). What is a Speech Act? In *Philosophy in America* (ed. M. Black), pp. 221-239. Cornell Univ. Press.

SHAHAR, Y., MIKSCH, S. & JOHNSON, P. (1998). The Asgaard Project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*, 29-51.

SHAHAR, Y., SHALOM, E., MAYAFFIT, A., YOUNG, O., GALPERIN, M., MARTINS, S. & AL., E. (2003). A distributed, collaborative, structuring model for a clinical-guideline digital-library. In *Proceedings of the Amer Med Informatics Assoc* (ed. M. MA), pp. 589-593, Washington, DC.

SHIFFMAN, R. (1997). Representation of clinical practice guidelines in conventional and augmented decision tables. *J Am Med Inform Assoc*. **4**, 382-393.

SHIFFMAN, R., BRANDT, C., LIAW, Y. & CORB, G. (1999). A design model for computer-based guideline implementation based on information management services. *J Am Med Inform Assoc*. **6**, 99-103.

SHIFFMAN, R., DIXON, J., BRANDT, C., ESSAIHI, A., HSIAO, A., MICHEL, G. & O'CONNELL, R. (2005). The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Medical Informatics and Decision Making* **5**.

SHIFFMAN, R. & GREENES, R. (1994). Improving clinical guidelines with logic and decision-table techniques: application to hepatitis immunization recommendations. *Medical Decision Making* **14**, 247-343.

SHIFFMAN, R., KARRAS, B., AGRAWAL, A., CHEN, R., MARENCO, L. & NATH, S. (2000). GEM: A proposal for a more comprehensive guideline document model using XML. *J Am Med Informatics Assoc* **7**, 488-498.

SHIFFMAN, R. & MICHEL, G. (2004). Toward Improved Guideline Quality: Using the COGS Statement with GEM. In *IMLA* (ed. M. F. e. al.), pp. 159-163. Amsterdam: IOS Press, San Francisco.

SHIFFMAN, R., MICHEL, G. & ESSAIHI, A. (2004a). Bridging the guideline implementation gap: a systematic approach to document-centered guideline implementation. *J Am Med Inform Assoc* **11**, 418-426.

SHIFFMAN, R., MICHEL, G., ESSAIHI, A. & MARCY, T. (2004b). Using a guideline-centered approach for the design of a clinical decision support system to promote smoking cessation. In *Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004)*. (ed. M. S. Kaiser K, Tu S.), pp. 152-156. Washington, DC, IOS Press.

SHIFFMAN, R., SHEKELLE, P., OVERHAGE, J., SLUTSKY, J., GRIMSHAW, J. & DESHPANDE, A. (2003). Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann Intern Med*. **139**, 493-498.

- STOLTE, J., ASH, J. & CHIN, H. (1999). The dissemination of clinical practice guidelines over an intranet: an evaluation. In *Proc AMLA*, pp. 960-964.
- SUTTON, D. & FOX, J. (2003 Sep-Oct). The syntax and semantics of the PROforma guideline modeling language. *J Am Med Inform Assoc* **10**, 433-443.
- SVATEK, V., KROUPA, T. & RUZICKA, M. (2001). Guide-X – a Step-by-step, Markup-Based Approach to Guideline Formalisation. In *First European Workshop on Computer-based Support for Clinical Guidelines and Protocols*, pp. 97-114.
- SVATEK, V. & RUZICKA, M. (2003). Step-by-step mark-up of medical guideline documents. *International Journal of Medical Informatics* **70**, 329–335.
- TERENZIANI, P., MONTANI, S., BOTTRIGHI, A., TORCHIO, M., MOLINO, G. & CORRENDO, G. (2004). The GLARE Approach to Clinical Guidelines: Main Features. In *Proceedings of the Symposium on Computerized Clinical Guidelines and Protocols*, vol. 101 (ed. M. S. Kaiser K, Tu S.).
- TIERNEY, E. A. (1995). Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *J Am Med Inf Assoc* **2**, 316-322.
- TU, S. & MUSEN, M. (2001). Modeling data and knowledge in the EON Guideline Architecture. In *Medinfo*, vol. 10, pp. 280–284.
- VAN DER MAAS, A., TEN HOOPEN, A. & TER HOFSTEDE, A. (2001). Progress with formalization in medical informatics? *J Am Med Inform Assoc.* **8**, 126-130.
- VAN RIJSBERGEN, C. J. (1979). *Information Retrieval, 2nd edition*. Dept. of Computer Science, University of Glasgow, London.
- VOTRUBA, P., MIKSCH, S., SEYFANG, A. & KOSARA, R. (2004). Tracing the Formalization Steps of Textual Guidelines. In *Proceedings of Stud Health Technol Inform*, vol. 101, pp.172-176.
- WANG, D., PELEG, M., TU, S., BOXWALA, A., GREENES, R., PATEL, V. & SHORTLIFFE, E. (2002a). Representation Primitives, Process Models and Patient Data in Computer-Interpretable Clinical Practice Guidelines: A Literature Review of Guideline Representation Models. *International Journal of Medical Informatics* **68**, 59-70.
- WANG, D., PELEG, M., TU, S., BOXWALA, A., GREENES, R., PATEL, V. & SHORTLIFFE, E. (2002b). Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines: a literature review of guideline representation models. *Int J Med Inf.* **18**, 59-70.

WANG, D., PELEG, M., TU, S., SHORTLIFFE, E. & GREENES, R. (2001). Representation of Clinical Practice Guidelines For Computer-Based Implementations. In *Medinfo*, vol. 10, pp. 285-289.

WANG, T., MORT, E., NORDBERG, P. & ET, A. (2003). Shared practice guidelines and staff education may reduce investigations without affecting clinical outcomes in the coronary care unit. *Evidence-based Cardiovascular Medicine* 7, 45-46.

WEIR, C. & SILK, B. (1992). Paramax Systems Corporation: description of the Paramax system used for MUC-4. In *The 4th conference on message understanding*, pp. 242-252.

YANGARBER, R. & GRISHMAN, R. (1998). NYU: Description of the Proteus /PET system as used for MUC-7 ST. In *MUC-7: Seventh Message Understanding Conference*, Columbia, MD.

Annexes

ANNEXE 1 – PUBLICATIONS RÉALISÉES DURANT LA THÈSE _____ 194

ANNEXE 2 – GUIDES DE BONNES PRATIQUES UTILISÉS POUR L'ANALYSE
LINGUISTIQUE _____ 195

ANNEXE 3 – GUIDES DE BONNES PRATIQUES UTILISÉS POUR ÉTABLIR LA
GRAMMAIRE _____ 198

ANNEXE 4 – EXEMPLE D'UN GUIDE DE BONNES PRATIQUES ANALYSÉ PAR G-DEE 200

ANNEXE 5 – RÉSULTATS D'ÉVALUATION DES EXPERTS AVEC LEUR APPRÉCIATION
POUR LE GUIDE DE BONNES PRATIQUES DE L'AVC _____ 209

Annexe 1 – Publications réalisées durant la thèse

Chapitre de livre

1 Gersende Georg, [Computerization of Clinical Guidelines: an Application of Medical Document Processing](#). In: BG Silverman, A Jain, A Ichalkaranje, LC Jain (Eds.). *Intelligent Paradigms for Healthcare Enterprises Systems Thinking*. Series: Studies in Fuzziness and Soft Computing, Springer-Verlag. 2005;184:1-30.

Article de revues internationales

2 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Extending the GEM Model to Support Knowledge Extraction from Textual Guidelines](#). *International Journal of Medical Informatics* 2005;74(2-4):79-87.

Actes de congrès internationaux avec comité de lecture

3 Gersende Georg, Marie-Christine Jaulent. [An Environment for Document Engineering of Clinical Guidelines](#). In: CP Friedman, editor. *Proceedings of the American Medical Informatics Association*; 22-26 October 2005; Washington, DC;2005;:276-280.

4 Gersende Georg, Isabelle Colombet, Marie-Christine Jaulent. [Structuring Clinical Guidelines through the Recognition of Deontic Operators](#). In: R Engelbrecht, editor. *Proceedings of Medical Informatics Europe 2005*; 28 August - 1 September 2005; Geneva, Switzerland; IOS Press Amsterdam. 2005;116:151-156.

5 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Synthesis of Elementary Single-Disease Recommendations to Support Guideline-Based Therapeutic Decision for Complex Polypathological Patients](#). In: M Fieschi, editor. *Proceedings of MEDINFO Symposium*; 7-11 September 2004; San Francisco; 2004;:38-42.

6 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Does GEM-Encoding Clinical Practice Guidelines Improve the Quality of Knowledge Bases? A Study with the Rule-Based Formalism](#). In: M Musen, editor. *Proceedings of the American Medical Informatics Association*; 8-12 November 2003; Washington, DC; 2003;:254-258.

7 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Managing Theoretical Single-Disease Guideline Recommendations for Actual Multiple-Disease Patients](#). In: M Dojat, editor. *Proceedings of 9th Conference on Artificial Intelligence in Medicine in Europe*; 18-22 October 2003; Protaras, Cyprus; Lecture Notes in Artificial Intelligence 2780, Springer-Verlag Berlin Heidelberg; 2003;:168-172.

8 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Interpretative Framework of Chronic Disease Management to Guide Textual Guideline GEM-Encoding](#). In: R Baud, editor. *Proceedings of Medical Informatics Europe*; 4-7 May 2003; Saint-Malo, France; IOS Press Amsterdam. 2003;95:531-536.

Actes de congrès nationaux avec comité de lecture

9 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Dérivation d'une base de connaissances à partir d'une instance GEM d'un guide de bonnes pratiques médicales textuel](#). In: R Dieng-Kuntz, editor. *14e journées francophones d'Ingénierie des Connaissances*; 1-4 July 2003; Laval, France; Presses Universitaires de Grenoble; 2003;:51-66. *In French*.

10 Gersende Georg, Brigitte Séroussi, Jacques Bouaud. [Synthèse de recommandations établies par pathologie pour l'aide à la thérapeutique des patients polypathologiques](#). *Xèmes Journées Françaises d'Informatique Médicale*; 4-5 September 2003; Tunis, Tunisie. *In French*.

Annexe 2 – Guides de Bonnes Pratiques utilisés pour l'analyse linguistique

- 1) Prise en charge diagnostique et traitement immédiat de l'accident ischémique transitoire de l'adulte - Mai 2004, ANAES (12 pages).
- 2) Évaluation du pronostic de l'insuffisance coronaire stable et modalités de suivi en dehors du traitement - Mars 2000, ANAES (11 pages).
- 3) Éducation thérapeutique du patient asthmatique adulte et adolescent - Juin 2001, ANAES (23 pages).
- 4) Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral – Aspects médicaux – Septembre 2002, ANAES (20 pages).
- 5) Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral – Aspects paramédicaux - Juin 2002, ANAES (29 pages).
- 6) Retour au domicile des patients adultes atteints d'accident vasculaire cérébral – Stratégies et organisation – Décembre 2003, ANAES (11 pages).
- 7) Le cancer du sein - Novembre 1998, ANAES (31 pages).
- 8) Éléments d'information des hommes envisageant la réalisation d'un dépistage individuel du cancer de la prostate - Document à l'usage des professionnels de santé - Septembre 2004, ANAES (22 pages).
- 9) Prise en charge diagnostique et thérapeutique du carcinome basocellulaire de l'adulte - Mars 2004, ANAES (24 pages).
- 10) CCQ (Céphalées chroniques quotidiennes) : Diagnostic, Rôle de l'abus médicamenteux, prise en charge - Septembre 2004, ANAES (17 pages).
- 11) Stratégies de choix des méthodes contraceptives chez la femme - Décembre 2004, ANAES (48 pages).
- 12) Évaluation diagnostique de la dénutrition protéino-énergétique des adultes hospitalisés - Septembre 2003, ANAES (15 pages).
- 13) Suivi du patient diabétique de type 2 à l'exclusion du suivi des complications - Janvier 1999, ANAES (14 pages).
- 14) Évaluation et stratégies de prise en charge de la douleur aiguë en ambulatoire chez l'enfant de 1 mois à 15 mois - Mars 2000, ANAES (26 pages).
- 15) Évaluation et prise en charge thérapeutiques de la douleur chez les personnes âgées ayant des troubles de la communication verbale - Octobre 2000, ANAES (18 pages).
- 16) Indication à visée diagnostique de l'endoscopie digestive haute en pathologie œso-gastro-duodénale de l'adulte à l'exclusion de l'échoendoscopie et l'entéroscopie – Mars 2001, ANAES (7 pages).
- 17) Éducation thérapeutique de l'enfant asthmatique - Juin 2002, ANAES (8 pages).

- 18) Prise en charge d'un épisode dépressif isolé de l'adulte en ambulatoire - Mai 2002, ANAES (13 pages).
- 19) Conduite à tenir devant une patiente ayant un frottis cervico-utérin anormal – Actualisation 2002, ANAES (22 pages).
- 20) Diagnostic et suivi virologiques des hépatites virales (à l'exclusion du dépistage en cas de dons de sang, d'organes ou de tissus) – Février 2001, ANAES (12 pages).
- 21) Modalités de prise de décision concernant l'indication en urgence d'une hospitalisation sans consentement d'une personne présentant des troubles mentaux - Avril 2005, HAS (18 pages).
- 22) Hémorragies du post-partum immédiat - Novembre 2004, HAS (12 pages).
- 23) Prise en charge des patients adultes atteints d'hypertension artérielle essentielle – Actualisation 2005, HAS (29 pages).
- 24) Prise en charge des patients adultes atteints d'hypertension artérielle essentielle – Recommandations cliniques et données économiques - Avril 2000, ANAES (17 pages).
- 25) Diagnostic et surveillance biologiques de l'hyperthyroïdie de l'adulte – Février 2000, ANAES (8 pages).
- 26) Prise en charge diagnostique et thérapeutique de l'hypertrophie bénigne de la prostate – Mars 2003, ANAES (13 pages).
- 27) Prise en charge de l'incontinence urinaire de la femme en médecine générale - Mai 2003, ANAES (17 pages).
- 28) Diagnostic et traitement curatif de l'infection bactérienne précoce du nouveau-né – Septembre 2002, ANAES (17 pages).
- 29) Diagnostic de l'insuffisance rénale chronique chez l'adulte – Septembre 2002, ANAES (21 pages).
- 30) Moyens thérapeutiques pour ralentir la progression de l'insuffisance rénale chronique chez l'adulte Recommandations Septembre 2004, ANAES (13 pages).
- 31) Orientations diagnostiques et prise en charge, au décours d'une intoxication éthylique aiguë, des patients admis aux urgences des établissements de soins – Septembre 2001, ANAES (25 pages).
- 32) Diagnostic, prise en charge et suivi des malades atteints de lombalgie chronique - Décembre 2000, ANAES (13 pages).
- 33) Prise en charge diagnostique et thérapeutique des lombalgies et lombosciatiques communes de moins de trois mois d'évolution – Février 2000, ANAES (6 pages).
- 34) Prise en charge thérapeutique des métastases hépatiques des cancers colo-rectaux – Janvier 2003, ANAES (18 pages).

- 35) Prise en charge diagnostique et thérapeutique de la migraine chez l'adulte et chez l'enfant : aspects cliniques et économiques - Octobre 2002, ANAES (21 pages).
- 36) Prise en charge de l'obésité de l'enfant et de l'adolescent - Septembre 2003, ANAES (14 pages).
- 37) L'ostéoporose chez les femmes ménopausées et chez les sujets traités par corticoïdes : méthodes diagnostiques et indications - Avril 2001, ANAES (7 pages).
- 38) Parodontopathies : diagnostic et traitements - Mai 2002, ANAES (19 pages).
- 39) Recommandations pour le suivi médical des patients asthmatiques adultes et adolescents - Septembre 2004, ANAES (18 pages) .
- 40) Rééducation dans le cadre du post-partum – Décembre 2002, ANAES (20 pages).
- 41) Prévention anténatale du risque infectieux bactérien néonatal précoce - Septembre 2001, ANAES (10 pages).
- 42) Stratégie de prise en charge du patient diabétique de type 2 à l'exclusion de la prise en charge des complications - Mars 2000, ANAES (21 pages).
- 43) Recommandations pratiques pour le diagnostic de la maladie d'Alzheimer – Février 2000, ANAES (8 pages).
- 44) Modalités de prise en charge de l'adulte nécessitant des soins palliatifs - Décembre 2002, ANAES (30 pages).
- 45) Les critères d'aboutissement du traitement d'orthopédie dento-faciale – Décembre 2003, ANAES (8 pages).
- 46) Prise en charge des traumatisés crâniens graves à la phase précoce - Janvier 1998, ANAES (17 pages).
- 47) Conduite à tenir lors de la découverte anténatale d'une ventriculomégalie cérébrale - Janvier 2004, ANAES (16 pages).

Annexe 3 – Guides de Bonnes Pratiques utilisés pour établir la grammaire

L'ensemble des documents suivants correspond au corpus médical de la faculté de médecine de Grenoble¹¹² :

- 1) Risque cardio-vasculaire et diabète (233d) – Docteur Pierre-Yves BENHAMOU – Novembre 2002 (7 pages).
- 2) L'insulinothérapie (233e) – Docteur Pierre-Yves BENHAMOU – Novembre 2002 (10 pages).
- 3) Hypertension artérielle essentielle (130a) – Docteur Jean-Philippe BAGUET, Professeur Jean-Michel MALLION – Août 2002 (9 pages).
- 4) Hypertension artérielle secondaire (130b) – Professeur Jean-Michel MALLION, Docteur Jean-Philippe BAGUET – Août 2002 (8 pages).
- 5) Insuffisance rénale aiguë (252) – Docteur Bénédicte JANBON, Professeur Philippe ZAOUI – Février 2004 (7 pages).
- 6) Le diabète de type 2 ou diabète non insulino-dépendant (DND) (223b) – Professeur Serge HALIMI – Avril 2003 (11 pages).

Documents issus du site de la Haute Autorité de Santé (HAS)¹¹³ :

- 1) Diagnostic, prise en charge et suivi des malades atteints de lombalgie chronique – Décembre 2000 (22 pages).
- 2) Prise en charge diagnostique et thérapeutique de la migraine chez l'adulte et chez l'enfant : aspects cliniques et économiques – Recommandations Octobre 2002 (21 pages).
- 3) Recommandations et références médicales – Supplément au n°41 du 7 décembre 1996 – ANAES (18 pages)
- 4) Évaluation des pratiques professionnelles dans les établissements de santé – les pneumonies aiguës communautaires – Décembre 2001 (ANAES) (55 pages).
- 5) Suivi du patient diabétique de type 2 à l'exclusion du suivi des complications – Janvier 1999 (14 pages).
- 6) Diagnostic et traitement de l'hypertension artérielle essentielle de l'adulte de 20 à 80 ans. Septembre 1997 (37 pages).
- 7) Prise en charge des patients adultes atteints d'hypertension artérielle essentielle – Avril 2000 (17 pages).

¹¹² <http://www.sante.ujf-grenoble.fr/sante/corpus/> (dernier accès le 10 juin 2006).

¹¹³ www.has-sante.fr/ (dernier accès le 10 juin 2006).

Conférences de consensus issus du site de la Haute Autorité de Santé (HAS)¹¹³ :

- 1) La maladie de Parkinson : critères diagnostiques et thérapeutiques – 3 mars 2000 – Amphithéâtre Charcot – Hôpital de la Pitié Salpêtrière – Paris (ANAES – Fédération française de neurologie) (17 pages).
- 2) La sclérose en plaques – Jeudi 7 et vendredi 8 juin 2001 – Amphithéâtre Charcot – Hôpital de la Pitié-Salpêtrière – Paris (ANAES – Fédération française de neurologie) (22 pages).
- 3) Stratégies thérapeutiques pour les personnes dépendantes des opiacés : place des traitements de substitution – 23 et 24 juin 2004 – Lyon (Étude normale supérieure) – (ANAES – Fédération française d'addictologie) (16 pages).
- 4) Hépatite C : Dépistage et traitement – Jeudi 16 et vendredi 17 janvier 1997 - Cité des Sciences et de l'Industrie – La Villette – Paris (ANAES, Association Française pour l'Étude du Foie, Société Nationale Française de Gastro-Entérologie) (31 pages).
- 5) Traitement de l'hépatite C – Mercredi 27 et jeudi 28 février 2002 – Maison de la Chimie – Paris (23 pages).
- 6) 8^e Conférence de consensus de la société francophone d'urgences médicales « prise en charge des coliques néphrétiques de l'adulte dans les services d'accueil et d'urgences » - 23 avril 1999 – Marseille (La Société Francophone d'Urgences Médicales) (33 pages).
- 7) Prise en charge des épilepsies partielles pharmacorésistantes – Mercredi 3 et jeudi 4 mars 2004 – Maison de l'Unesco – Paris (Société française de neurologie) (14 pages).
- 8) La sclérose en plaques – Jeudi 7 et vendredi 8 juin 2001 – Amphithéâtre Charcot – Hôpital de la Pitié-Salpêtrière – Paris (Société Française de Neurologie) (22 pages).
- 9) Prise en charge de l'urticaire chronique – Mercredi 8 janvier 2003 – Institut Pasteur – Paris (Société Française de Dermatologie) (19 pages).

**Prise en charge initiale des patients adultes atteints d'accident vasculaire cérébral - Aspects médicaux.
(Septembre 2002)¹¹³**

RECOMMANDATIONS

I. INTRODUCTION

I.1. Thème des recommandations

Les recommandations concernent la prise en charge thérapeutique des patients atteints d'accident vasculaire cérébral (AVC) à la phase aiguë, c'est-à-dire environ dans les 15 premiers jours, à l'exclusion de l'hémorragie méningée (1). Elles s'adressent à tous les professionnels de santé impliqués à cette phase de la prise en charge (2).

I.2. Niveau de preuve des recommandations

Les recommandations sont classées en grade A, B ou C selon les modalités suivantes :

- une recommandation de grade A est fondée sur une preuve scientifique établie par des études de fort niveau de preuve, par exemple essais comparatifs randomisés de forte puissance et sans biais majeur, méta-analyse d'essais randomisés, analyse de décision basée sur des études bien menées ;
- une recommandation de grade B est fondée sur une présomption scientifique fournie par des études de niveau intermédiaire de preuve, par exemple essais comparatifs randomisés de faible puissance, méta-analyse de méthodologie critiquable, études comparatives non randomisées bien menées, études de cohorte ;
- une recommandation de grade C est fondée sur des études de moindre niveau de preuve, par exemple études cas-témoins, séries de cas (3).

En l'absence de précision, les recommandations proposées reposent sur un accord professionnel au sein du groupe de travail et du groupe de lecture (4).

I.3. Définitions et repères épidémiologiques

Les AVC sont un groupe d'affections au sein desquelles il faut distinguer :

- les AVC ischémiques ou infarctus cérébraux (85 % des AVC) ;
- les AVC hémorragiques (15 %), subdivisés en hémorragies cérébrales (10 %) et méningées (5 %) (5).

Le terme "accident ischémique cérébral" (AIC) regroupe les infarctus cérébraux et les accidents ischémiques transitoires (AIT) qui correspondent à une ischémie réversible dont les symptômes régressent en moins de 24 heures (6). Les AIT sont le meilleur signe avant-

coureur de l'infarctus cérébral dont ils partagent les mécanismes et les causes (7). L'AIT impose une enquête étiologique rapide en vue d'un traitement adapté (8).

Dans les pays industrialisés, les AVC représentent la 3^e cause de mortalité après les cardiopathies et les cancers, et la 1^{re} cause de handicap non traumatique de l'adulte (9). En France, 169 843 AVC ischémiques et hémorragiques, incluant les AIT, ont été recensés par le PMSI durant l'année 1999 (10).

I.4. Diagnostic de l'AVC, de sa nature et de son territoire

Le diagnostic d'AVC, évoqué devant des troubles neurologiques focaux ou des troubles de la vigilance d'installation soudaine, **doit être confirmé** par l'imagerie cérébrale :

- l'imagerie cérébrale (IRM, scanner) est nécessaire sans délai pour affirmer le diagnostic d'AVC et préciser sa nature ischémique ou hémorragique. Actuellement, l'IRM-séquences écho-planar (méthode permettant d'obtenir une coupe IRM dans un temps d'acquisition de l'ordre de 100 ms) est l'examen le plus performant car elle permet le diagnostic très précoce (dès la 1^{re} heure) de l'ischémie et de l'hémorragie cérébrales, ainsi que l'évaluation du caractère récent ou ancien, de l'étendue et de la sévérité des lésions. Elle permet, en outre, l'étude des vaisseaux.
- le scanner cérébral sans injection de produit de contraste permet le diagnostic en urgence d'hémorragie cérébrale. Les signes précoces d'ischémie cérébrale sont inconstants et d'interprétation difficile.
- l'IRM avec séquences d'angioRM veineuse est l'examen non invasif le plus performant pour le diagnostic de thrombose veineuse cérébrale (11).

Le diagnostic étiologique est une étape fondamentale qui conditionne à la fois la prise en charge immédiate et la prévention secondaire (12).

II. RECOMMANDATIONS POUR LA PRISE EN CHARGE THERAPEUTIQUE DES PATIENTS AYANT UN AVC.

L'AVC est une urgence diagnostique et thérapeutique (13).

II.1. Surveillance neurologique et des paramètres vitaux

II.1.1. Surveillance neurologique

L'état de vigilance et l'état neurologique **doivent être notés** le plus tôt possible par le médecin qui prend en charge le patient (14).

L'évolution de la vigilance et de l'état neurologique (jusqu'à stabilisation du déficit) **doit être surveillée** à un rythme déterminé par prescription médicale (en fonction de l'état du patient), et transcrite dans le dossier du patient (15).

L'utilisation d'échelles du type suivant **est recommandée** :

- pour l'état de vigilance, une échelle simple adaptée de celle d'Orgogozo : vigilance normale/éveil spontané, somnolence/obnubilation, réaction (éveil) à l'ordre verbal, stupeur/réaction à la douleur, coma/aucune réaction adaptée. Le score de Glasgow, plus adapté aux comas traumatiques qu'aux AVC, peut aussi être utilisé ;

- pour l'état neurologique, plusieurs échelles spécifiques à l'AVC sont utilisées, en langue française ou anglaise ou en traduction française : échelle d'Orgogozo, échelle canadienne, échelle scandinave, échelle du National Institutes of Health (NIH). L'échelle du NIH est l'échelle de référence en cas de fibrinolyse (16).

II.1.2. Surveillance des paramètres vitaux

La pression artérielle **doit être régulièrement surveillée** à un rythme déterminé par prescription médicale, de façon d'autant plus rapprochée que l'AVC est récent (17).

Un ECG **doit être réalisé** dès le début de la prise en charge (18). Une surveillance continue par cardioscope au cours des 48 premières heures de l'AVC permet de dépister une arythmie paroxystique ou une autre pathologie cardiaque associée (19).

La fonction respiratoire et la température corporelle **doivent être surveillées** à un rythme déterminé par prescription médicale (20).

II.2. Prise en charge des complications générales

II.2.1. Pression artérielle

Il est recommandé de respecter l'hypertension artérielle contemporaine de la phase aiguë d'un AVC ischémique (grade C) ou d'une hémorragie cérébrale, sauf dans le cas suivants :

- AVC ischémique.

1) Un traitement fibrinolytique est envisagé (30). Dans ce cas, la pression artérielle doit être abaissée en dessous de 185/110 mmHg avant de débiter le traitement fibrinolytique (accord professionnel à partir des critères d'inclusion dans les études de référence).

2) Un traitement fibrinolytique n'est pas envisagé :

- persistance d'une pression artérielle supérieure à 220/120 mmHg (accord professionnel) ;
- complication de l'HTA menaçante à court terme (ex. : dissection aortique, insuffisance cardiaque décompensée, encéphalopathie hypertensive).

- Hémorragie cérébrale.

Bien que certains proposent l'abaissement à 185/110 mmHg du seuil tensionnel d'intervention thérapeutique, il n'existe pas de preuve scientifique à l'appui de cette différence de seuil comparativement à l'AVC ischémique (21).

En cas de **traitement de l'hypertension artérielle, il est recommandé d'utiliser** préférentiellement la perfusion intraveineuse (IV) pour un ajustement tensionnel précis (22). Les voies intramusculaire et sublinguale **sont à éviter** (23). L'utilisation préférentielle de l'urapidil ou du labétalol ou de la nicardipine **est recommandée**, en évitant les doses de charge (accord professionnel) (24).

La pression artérielle **doit être abaissée progressivement et maintenue** en dessous de 220/120 mmHg, en surveillant l'état neurologique afin de dépister une aggravation du déficit (25). Il n'existe pas d'étude permettant de définir un objectif tensionnel précis (26).

Il est recommandé de maintenir le traitement antihypertenseur préexistant (27). Il n'existe pas d'étude permettant de définir un seuil tensionnel précis en dessous duquel un traitement antihypertenseur préexistant **devrait être arrêté** (28).

II.2.2. Troubles respiratoires

La liberté des voies aériennes supérieures **doit être assurée** (29). L'encombrement bronchique et les pneumopathies d'inhalation **doivent être prévenues** (30). L'oxygénothérapie systématique **n'est pas recommandée** (grade B) (31).

II.2.3. Hyperthermie

En l'absence d'étude définissant un seuil d'intervention, **il est recommandé de traiter** une hyperthermie au-dessus de 37,5 °C par un antipyrétique type paracétamol (accord professionnel) (32). Les infections documentées **doivent être traitées** par une antibiothérapie adaptée (33).

II.2.4. Troubles de la déglutition et nutrition

Les troubles de la déglutition **doivent être recherchés** systématiquement avant la première alimentation, car ils exposent au risque de pneumopathie d'inhalation (grade B) (34). **En cas de troubles de la déglutition, l'alimentation orale doit être interrompue et remplacée** par une alimentation par sonde naso-gastrique, adaptée à l'état nutritionnel (35).

En l'absence d'alimentation orale, un apport calorique adéquat sera fourni par voie entérale, à partir de la 48^e heure (accord professionnel) (36).

La place de la gastrostomie d'alimentation à la phase aiguë n'a pas été évaluée (37).

II.2.5. Troubles hydro-électrolytiques, hyperglycémie

L'équilibre hydro-électrolytique et glycémique **doit être surveillé et les anomalies corrigées** (38). Si **une perfusion IV est nécessaire, il est recommandé d'utiliser** du sérum physiologique plutôt que du soluté glucosé (39).

En l'absence d'étude définissant un seuil d'intervention, **il est recommandé de traiter** par insulinothérapie les patients dont la glycémie est ≥ 10 mmol/l (accord professionnel) (40).

II.2.6. Prévention des complications thrombo-emboliques veineuses

Le lever précoce **est recommandé**, sauf en cas de **situation hémodynamique cérébrale précaire se manifestant par une aggravation à l'orthostatisme** (41).

Un traitement préventif des complications thrombo-emboliques veineuses par héparine de bas poids moléculaire (HBPM) à faibles doses **est recommandé**, dès les 24 premières heures, chez les patients ayant un AVC ischémique aigu responsable d'une immobilisation ou d'un déficit du membre inférieur (accord professionnel) (42). Cette indication **doit cependant être discutée** en fonction du **risque hémorragique intra et extracrânien** (43).

Il n'existe pas de preuves suffisantes pour recommander, en routine, la contention des membres inférieurs en prévention des complications thrombo-emboliques veineuses chez les patients ayant un AVC ischémique (44). Elle **peut néanmoins être recommandée** en cas de **contre-indication à l'utilisation des héparines** (45).

La sécurité d'utilisation de l'héparine étant mal connue chez les patients ayant un AVC hémorragique, l'utilisation d'une contention élastique immédiate des membres inférieurs, éventuellement suivie d'une héparinothérapie à doses préventives après 24-48 heures, **est recommandée**, bien que cette attitude ne repose sur aucune évaluation (46).

II.2.7. Prévention de l'hémorragie digestive

En l'absence d'étude spécifique concernant l'AVC, l'utilisation à titre préventif d'antiulcéreux **n'est pas recommandée**, sauf en cas d'**antécédent ulcéreux** (47).

II.2.8. Troubles anxio-dépressifs

Le traitement des troubles anxio-dépressifs n'a pas été suffisamment évalué à la phase aiguë de l'AVC (48).

II.2.9. Autres mesures thérapeutiques

Les recommandations relatives à la kinésithérapie, l'orthophonie, le nursing, la prévention des complications intestinales, la prise en charge de l'incontinence et de la rétention d'urine, la prévention des complications cutanées et la prise en charge des troubles anxio-dépressifs **sont traitées** dans les recommandations pour la pratique clinique sur la "prise en charge pluridisciplinaire initiale des patients adultes atteints d'AVC - aspects paramédicaux" (49).

II.3. Prise en charge des complications neurologiques de l'AVC

II.3.1. Œdème cérébral

Il faut éviter toute hyperhydratation, notamment intracellulaire (50). Si **une restriction hydrique est indiquée**, elle **doit être modérée**, de l'ordre de 1 000 ml/j (51).

Bien qu'il n'existe pas de preuve scientifique de leur efficacité, les agents hyperosmolaires (ex : mannitol, glycérol) **peuvent être utilisés** pendant moins de 5 jours chez les patients dont l'état clinique se détériore du fait d'un œdème cérébral, notamment en cas de **signes d'engagement cérébral** (52).

Les corticostéroïdes **ne doivent pas être utilisés** en cas d'**œdème cérébral (grade B)**, car ils sont inefficaces et dangereux (53).

Le traitement des AVC ischémiques massifs par chirurgie de décompression (cf II 6 2) ou par hypothermie externe est en cours d'évaluation (54). L'hypothermie **ne peut être actuellement recommandée** (55). Le bénéfice des techniques de réanimation hyperventilation mécanique, thiopental) n'est pas démontré (56).

II.3.2. Épilepsie

Le traitement antiépileptique **n'est pas recommandé** en prophylaxie des crises (57). En cas de **crise à la phase aiguë**, un traitement antiépileptique **est recommandé** pour prévenir les récurrences de crises (58). Les différents médicaments n'ont pas été évalués spécifiquement dans le contexte de l'AVC, y compris en cas d'état de mal (59). La durée optimale du traitement n'a pas été évaluée (60). **Il n'y a pas d'argument pour proposer** un traitement au long cours (61).

II.4. Traitement de l'AVC ischémique artériel

II.4.1. Antithrombotiques

Un traitement antiplaquettaire par aspirine (à la posologie de 160 à 300 mg/j) **est recommandé** dès que possible après un AVC ischémique (grade A), sauf si **un traitement thrombolytique est envisagé** (62).

L'utilisation systématique d'héparine (héparine non fractionnée, héparines de bas poids moléculaire, ou héparinoïdes) à doses curatives **n'est pas recommandée** à la phase aiguë de l'AVC ischémique, y compris dans la fibrillation auriculaire non valvulaire (grade A) (63).

Les essais thérapeutiques n'ayant pas ciblé toutes les situations étiologiques, l'héparine à doses curatives **peut être utilisée** dans des indications sélectives, présumées à haut risque de récurrence ou d'extension, telles que les cardiopathies à haut risque embolique, les sténoses artérielles sévères, la présence d'un thrombus intraluminal ou les dissections artérielles extracrâniennes (accord professionnel) (64). Le bénéfice potentiel **doit être mis en balance** avec le risque hémorragique cérébral qui est particulièrement élevé en cas d' **infarctus cérébral de grande taille responsable de troubles de la vigilance**, de signes précoces d'ischémie étendue à l'imagerie ou d'hypertension artérielle non contrôlée (65).

II.4.2. Fibrinolyse

Un traitement thrombolytique par rt-PA (altéplase) par voie veineuse **est recommandé** chez les patients ayant un AVC ischémique à condition que le traitement **puisse être instauré** dans les 3 heures suivant l'apparition des symptômes d'AVC et qu'il n'existe pas de contre-indications à ce traitement (voir *annexes 1 et 2*) (grade A) (66).

La dose de rt-PA **préconisée est de** 0,9 mg/kg, dose maximale 90 mg, 10 % en bolus et 90 % en perfusion d'une heure (grade A) (67).

Un traitement antiplaquettaire ou par héparine **est contre-indiqué** au cours des 24 heures suivant l'administration de rt-PA (68). Le rt-PA **peut être utilisé** chez les malades qui suivaient un traitement par l'aspirine au moment de l'AVC bien que le rapport bénéfice/risque soit incertain (69).

La pression artérielle **doit être** < 185/110 mmHg avant de débiter le traitement et maintenue à ce niveau pendant le traitement et durant les 24 heures suivantes (70).

Le traitement **doit être administré** uniquement par des médecins formés et expérimentés en neurologie (71).

Les essais cliniques ayant prouvé l'efficacité du rt-PA IV ont tous été réalisés dans des centres spécialisés dans la prise en charge des AVC aigus (72). Les données disponibles (séries de cas) dans les centres non spécialisés ont montré une surmortalité par hémorragie cérébrale par rapport aux résultats observés dans les centres spécialisés, notamment du fait du non-respect des indications (73). **Il est donc recommandé de ne pas utiliser** le rt-PA IV en dehors d'une structure spécialisée dans la prise en charge des AVC (accord professionnel) (74).

La streptokinase IV **ne doit pas être utilisée** dans la fibrinolyse des AVC ischémiques (grade A) (75).

La seule étude randomisée évaluant le bénéfice de la fibrinolyse intra-artérielle a montré l'efficacité de la r-pro-urokinase associée à l'héparine chez les patients ayant un AVC ischémique de moins de 6 heures, lié à une occlusion de l'artère cérébrale moyenne démontrée par l'angiographie cérébrale (76). **Des recommandations précises ne peuvent pas être faites** actuellement concernant cette technique (77).

II.4.3. Autres traitements

Les médicaments à visée neuroprotectrice **ne sont pas recommandés**, car aucune étude n'a démontré leur efficacité (grade A) (78). Une évaluation de ces produits dans les 3 heures après l'AVC **devrait être réalisée** (79).

L'hémodilution **n'est pas recommandée** dans le traitement de l'AVC ischémique aigu (grade B) (80).

II.5. Traitement des thromboses veineuses cérébrales

Un traitement anticoagulant à dose curative **est recommandé** chez tous les patients ayant une thrombose veineuse cérébrale certaine, y compris en cas d'**aspect hémorragique à l'imagerie (grade B)** (81).

II.6. Indications du traitement neurochirurgical

Les indications du traitement neurochirurgical des AVC ischémiques ou hémorragiques **sont insuffisamment évaluées** (82).

II.6.1. Hémorragies cérébrales

La décision de traitement neurochirurgical repose sur :

- des critères cliniques : âge du patient, traitement en cours, en particulier anticoagulant, niveau de vigilance (score de Glasgow), taille pupillaire, évolutivité ;
 - des critères neuroradiologiques : taille et siège de l'hématome, hémorragie sous-arachnoïdienne associée, effet de masse, taille des ventricules. L'appréciation de ces signes demande une expertise, au mieux un transfert d'images vers le service de neurochirurgie (83).
- Hématome du cervelet
 - L'hydrocéphalie par obstruction du 4e ventricule est un argument d'intervention (dérivation ventriculaire externe ou ventriculo-cisternostomie endoscopique) chez les

patients dont l'état clinique s'aggrave, vigilants ou avec troubles modérés de la vigilance et en l'absence de signes de compression du tronc cérébral (grade C) (84).

- L'évacuation de l'hématome **peut être envisagée** dans les cas suivants : hématome > 3 cm avec hydrocéphalie (grade C), coma (de durée inférieure à 2 heures si **le coma est profond**) ou **compression évolutive du tronc cérébral** (85).

- Hémorragie cérébrale lobaire avec aggravation clinique en l'absence de contreindication liée à l'état général (grade B) (86).

Il n'y a pas d'indication neurochirurgicale :

- en cas d'hématomes de petite taille (< 10 cm³) ou de déficit neurologique mineur (grade B) ;
- chez les patients ayant un hématome hémisphérique et un score de Glasgow = 4 (grade B) (87).

Pour tous les autres cas d'hémorragie cérébrale, la meilleure thérapeutique reste imprécise (88).

II.6.2. AVC ischémiques

Une dérivation ventriculaire externe transitoire du LCR **est indiquée** chez les patients ayant un infarctus cérébelleux avec hydrocéphalie aiguë (89). Une exérèse du territoire infarcté **est envisagée** de façon exceptionnelle en cas de **persistance de compression symptomatique du tronc cérébral après dérivation, si l'IRM ne montre pas d'extension de l'ischémie au tronc cérébral lui-même** (90).

La craniectomie de décompression des infarctus sylviens malins est en cours d'évaluation (91). En attendant des preuves scientifiques, ce traitement **peut être envisagé** chez le sujet jeune atteint d'un infarctus sylvien récent étendu oedémateux (grade C) (92).

II.7. Indications de la prise en charge en réanimation médicale

Les indications de la prise en charge en réanimation sont rares :

- traitement des comorbidités sévères curables chez les patients ayant un bon pronostic neurologique, telles que pneumopathie de déglutition, embolie pulmonaire ;
- hypertension intracrânienne si un geste chirurgical est envisageable. En cas d'engagement cérébral, la ventilation devrait être réservée aux situations dans lesquelles on peut envisager un geste chirurgical rapide.
- situations neurologiques instables et réversibles telles que thromboses veineuses cérébrales avec troubles de conscience, état de mal épileptique, éclampsie (93).

La décision d'instituer ou non une ventilation mécanique dans les suites d'un AVC aigu **doit dans la mesure du possible être concertée** entre réanimateurs, urgentistes, neurologues et famille du patient (94). L'absence de réflexes photomoteur et cornéen est un indicateur de forte probabilité de décès à prendre en compte dans la décision de ventilation mécanique (95).

III. UNITES NEURO-VASCULAIRES.

Bien qu'il n'existe pas de définition univoque des unités neuro-vasculaires, le groupe considère, en accord avec l'expérience internationale, que les caractéristiques suivantes sont importantes :

- une prise en charge globale des patients intégrant le diagnostic et les soins à la phase aiguë, le traitement des complications, la réadaptation et la prévention des événements vasculaires ;
- une équipe multiprofessionnelle spécialisée en pathologie neuro-vasculaire, coordonnée et régulièrement formée ;
- un site dédié à la prise en charge initiale de ces patients et disposant d'un plateau technique approprié ;
- une organisation de la filière de soins en amont et en aval ;
- la prise en compte des objectifs du patient et l'implication de la famille dans la prise en charge (96).

Chaque fois que possible, **il est recommandé d'hospitaliser** les patients suspects d'AVC en unités neuro-vasculaires (grade B), hormis les patients justifiant le recours d'emblée à un service de réanimation médicale ou de neurochirurgie (97).

Le groupe considère qu'il est urgent d'offrir aux patients atteints d'AVC une prise en charge structurée intégrant les soins à la phase aiguë et les soins de suite et de réadaptation (98).

Une amélioration de la prise en charge préhospitalière, en particulier en termes de rapidité de prise en charge et d'orientation, et de la filière de soins en aval est indispensable pour que l'efficacité des unités neuro-vasculaires soit optimale (99).

IV. ORGANISATION DE LA FILIERE DES SOINS ? PRISE EN CHARGE PREHOSPITALIERE.

Le public **doit être sensibilisé** par une campagne d'information visant à lui apprendre à reconnaître les signes d'appel d'un AIT ou d'un AVC et à considérer ces signes comme une urgence (100). Tous les professionnels de santé concernés **doivent contribuer** à faire considérer l'AVC comme une urgence médicale (101).

Le transfert du patient à l'hôpital, idéalement directement dans une unité neuro-vasculaire ou dans un service d'accueil des urgences, **doit être** le plus rapide possible, car les résultats de la prise en charge en dépendent (102). L'appel du centre téléphonique des urgences médicales **est recommandé** (appel au centre 15) (103).

Annexe 5 – Résultats d'évaluation des experts avec leur appréciation pour le Guide de Bonnes Pratiques de l'AVC

	Expert 1	Expert 2	Expert 3	Commentaires
PHRASES QUI SONT CONSIDEREES COMME DES RECOMMANDATIONS PAR LES EXPERTS				
L'AIT impose une enquête étiologique rapide en vue d'un traitement adapté (8).	VN	VN	FN	<p>« Recommandation mal faite. Problème de situation dans le contexte. Ce n'est pas le sujet (à voir si elle existe ailleurs). Recommandation hors sujet. Néanmoins faire une reconnaissance du verbe « impose ».</p> <p>« Principe général. Phrase qui est l'occasion d'introduire une recommandation. C'est « rapide » qui influence et la présence de « impose ». On aurait pu écrire « il est recommandé dans le cadre de l'AIT ... ». Recommandation qui reste générale. C'est le fait de le faire rapidement qui implique une recommandation. »</p>
L'AVC est une urgence diagnostique et thérapeutique (18).	VN	VN	FN	<p>« Le fait de définir l'AVC comme une urgence. C'est l'urgence qui implique une recommandation. C'est plus le contenu de la phrase qui fait penser à une recommandation. »</p>
Une surveillance continue par cardioscope au cours des 48 premières heures de l'AVC permet de dépister une arythmie paroxystique ou une autre pathologie cardiaque associée (27).	VN	FN	VN	<p>« C'est une recommandation. »</p>
En l'absence d'alimentation orale, un apport calorique adéquat sera fourni par voie entérale, à partir de la 48 ^e heure (accord professionnel) (49).	FN	VN	FN	<p>« C'est comme si on disait : il faut fournir. »</p> <p>« Futur impératif »</p>
Le groupe considère qu'il est urgent d'offrir aux patients atteints d'AVC une prise en charge structurée intégrant les soins à la phase aiguë et les soins de suite et de réadaptation (117).	VN	VN	FN	<p>« On ne connaît pas la cible. Recommandation générale de Santé Publique. Recommandations pour les institutions et pour stimuler les cliniques. »</p>

Une amélioration de la prise en charge préhospitalière, en particulier en termes de rapidité de prise en charge et d'orientation, et de la filière de soins en aval est indispensable pour que l'efficacité des unités neuro-vasculaires soit optimale (118).	VN	VN	FN	« Même chose que précédemment (117). C'est une spécification de la précédente. »
PHRASES QUI NE SONT PAS CONSIDEREES COMME DES RECOMMANDATIONS PAR LES EXPERTS				
Il n'existe pas d'étude permettant de définir un seuil tensionnel précis en dessous duquel un traitement antihypertenseur préexistant devrait être arrêté (41).	FP	VP	FP	« Pas d'étude. Ce n'est pas une recommandation car c'est une affirmation. C'est un constat. Ca ne veut pas dire qu'il faille définir un seuil. » « Justification de la recommandation négative qui précède. C'est plus un argumentaire. »
Les recommandations relatives à la kinésithérapie, l'orthophonie, le nursing, la prévention des complications intestinales, la prise en charge de l'incontinence et de la rétention d'urine, la prévention des complications cutanées et la prise en charge des troubles anxio-dépressifs sont traitées dans les recommandations pour la pratique clinique sur la "prise en charge pluridisciplinaire initiale des patients adultes atteints d'AVC - aspects paramédicaux" (62).	FP	FP	VP	« C'est traité ailleurs. Ce n'est pas une recommandation. C'est de l'information. » « C'est une information donc ce n'est pas une recommandation. »
Le bénéfice potentiel doit être mis en balance avec le risque hémorragique cérébral qui est particulièrement élevé en cas d' infarctus cérébral de grande taille responsable de troubles de la vigilance , de signes précoces d'ischémie étendue à l'imagerie ou d'hypertension artérielle non contrôlée (78).	FP	VP	VP	« Présentation de données. Ca peut se discuter. Domaine du flou. » « Rapport bénéfice/risque. Ce n'est pas une recommandation. Suite de la précédente (77). C'est une assertion. »
Une évaluation de ces produits dans les 3 heures après l'AVC devrait être réalisée (92).	VP	FP	VP	« Ce n'est pas une recommandation. »
Les indications du traitement neurochirurgical des AVC ischémiques ou hémorragiques sont insuffisamment évaluées (95).	FP	FP	FP	« Ce n'est pas une recommandation. » « C'est un constat. » « Problème de l'argumentation négative. »