



**HAL**  
open science

## Analyse et structuration automatique des guides de bonnes pratiques cliniques : essai d'évaluation.

Amanda Bouffier, Catherine Duclos, Thierry Poibeau

### ► To cite this version:

Amanda Bouffier, Catherine Duclos, Thierry Poibeau. Analyse et structuration automatique des guides de bonnes pratiques cliniques : essai d'évaluation.. 19es Journées Francophones d'Ingénierie des Connaissances (IC 2008), Jun 2008, Nancy, France. pp.135-146. hal-00416694

**HAL Id: hal-00416694**

**<https://hal.science/hal-00416694>**

Submitted on 14 Sep 2009

**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.

# Analyse et structuration automatique des guides de bonnes pratiques cliniques : essai d'évaluation

Amanda Bouffier<sup>1</sup>, Catherine Duclos<sup>2</sup>, Thierry Poibeau<sup>1</sup>

<sup>1</sup>Laboratoire d'Informatique de Paris-Nord (LIPN),  
UMR CNRS 7030 et Université Paris 13  
{prenom.nom}@lipn.univ-paris13.fr

<sup>2</sup>Laboratoire d'Informatique Médicale et BioInformatique (EA 3969 - LIM&BIO),  
UFR SMBH, Université Paris 13  
catherine.duclos@avc.aphp.fr

**Résumé :** Les guides de bonnes pratiques cliniques (GBPC) sont des textes constitués de recommandations valides. Ils ont pour but de diffuser des synthèses de résultats démontrés et de normaliser des conduites à tenir dans des situations cliniques données. Cet article présente un outil appelé GemFrame, destiné à faciliter la consultation des GBPC en proposant de nouveaux modes d'accès à l'information sur support électronique. Pour ce faire, un travail d'analyse et de structuration des GBPC est nécessaire. Nous présentons le système GemFrame, permettant cette structuration reposant sur une analyse semi-automatisée. Le système vise à reconnaître automatiquement les segments « conditions » et les segments « incitation à l'action », puis à calculer la portée des conditions sur les actions. Nous présentons ici une évaluation détaillée sur plusieurs guides. Nous montrons d'abord l'intérêt de l'approche, puis nous détaillons le processus d'évaluation fondé sur la comparaison des résultats obtenus automatiquement avec ceux obtenus manuellement (suite à l'élaboration d'une « référence »).

**Mot clés :** Guides de bonnes pratiques cliniques, modélisation, analyse du discours

## 1 Introduction

Les guides de bonnes pratiques cliniques (GBPC) sont des textes constitués de recommandations valides. Ils ont pour but de diffuser des synthèses de résultats démontrés obtenus à partir d'essais cliniques et de normaliser des conduites à tenir dans des situations cliniques données. L'adhésion des médecins à ces guides doit conduire à une médecine de qualité basée sur des preuves scientifiques.

Bien que l'utilité de ces guides soit évidente, leur utilisation en pratique connaît un certain nombre d'obstacles. Parmi ces obstacles, il est possible d'identifier des

facteurs intrinsèques liés au médecin et à sa capacité à connaître, adhérer et croire au guide [Cabana et al 1999], mais aussi des facteurs liés au texte lui-même. En effet, ces textes, issus de consensus d'experts, présentent un niveau de complexité élevé (longueur, manque de standardisation, manque de clarté, construction complexe faisant intervenir des relations causales ou temporelles) [Patel *et al.* 2001, Elkin *et al.* 2000]. Ces textes peuvent également présenter des ambiguïtés [Codish 2005], des imprécisions, ou ne pas traiter l'ensemble des situations possibles [Shiffman, Greene 1994], [Shiffman *et al.* 1999],

L'amélioration de ces documents (en facilitant leur compréhension, clarifiant leur contenu et en contrôlant leur cohérence) semble pouvoir faciliter leur diffusion [Patel *et al.* 2001, Shiffman 2004, Shiffman 2003]. Des modèles de représentation des connaissances contenues dans les GBPC ainsi que des formats documentaires ont été proposés afin de structurer les guides, les rendre plus homogènes et plus facilement lisibles.

Cet article présente un outil appelé GemFrame, destiné à faciliter la structuration des guides par un processus semi-automatisé. Les GBPC étant des textes d'incitation à l'action, il sont principalement constitués de conditions, et des incitations à l'action dépendant de ces conditions. Le système GemFrame vise à reconnaître automatiquement les segments « conditions » et les « segments incitation à l'action », puis à calculer la portée des conditions sur les actions. Plutôt que d'insister sur le mode de fonctionnement de l'outil qui a été détaillé dans diverses publications [Bouffier et Poibeau, 2007; Bouffier, 2007], nous présentons ici une évaluation fondée sur l'analyse de plusieurs guides. Nous montrons d'abord l'intérêt de l'approche puis nous détaillons le processus d'évaluation fondé sur la comparaison des résultats obtenus automatiquement avec ceux obtenus manuellement (suite à l'élaboration d'une « référence »).

Nous revenons dans un premier temps sur l'état de l'art afin de montrer les principales réalisations dans ce cadre. Nous présentons ensuite rapidement le fonctionnement de GemFrame avant d'en venir à l'évaluation du système, effectuée automatiquement à partir d'une comparaison avec un découpage manuel. Nous terminons en discutant les résultats obtenus et leur pertinence pour la tâche.

## **2 Vers une modélisation automatique des GBPC**

Nous présentons dans cette partie différents formalismes permettant de rendre compte des GBPC, ainsi que les outils permettant une automatisation plus ou moins grande du processus de modélisation.

### **2.1 Intérêt de modéliser les GBPC**

Des modèles et des formalismes de représentation de connaissances ont été proposés pour structurer ces guides et les informatiser. L'informatisation permet d'envisager un certain nombre de fonctionnalités simplifiant la gestion du cycle de vie des GBPC. Ainsi, il peut être possible d'identifier l'évolution des connaissances

entre deux versions d'un guide et d'envisager la maintenance automatisée de la base de connaissances [Shahar 2004, Shalom 2005, Shiffman 2000]. Il est aussi possible de vérifier la cohérence du guide grâce à une élaboration automatique des arbres de décision et une recherche automatique des situations manquantes [Shiffman 1997]. Enfin l'informatisation peut permettre d'aider à l'écriture du guide et faciliter son intégration dans des systèmes d'aide à la décision.

L'intégration des guides de bonnes pratiques dans des systèmes d'aide à la décision est un moyen d'apporter au médecin, dans son environnement de travail, un accès à la connaissance médicale. Diverses modalités d'intégrations peuvent être envisagées. Par exemple, dans le cadre du projet ASTI (Aide à la Stratégie Thérapeutique Informatisée), deux modes d'utilisation des connaissances issues des guides de bonnes pratiques ont été envisagés : le premier est un mode guidé qui aide le médecin à naviguer dans l'arborescence du guide en fonction des caractéristiques de son patient pour aboutir à une décision (par exemple le traitement à prescrire) [Séroussi 2001]. Le mode critique constitue, lui, le deuxième mode d'utilisation des connaissances issues du GBPC. Il fonctionne en feedback d'une décision du médecin : le système confronte la décision thérapeutique du médecin à celle que le système trouve en utilisant les données d'historique thérapeutique, les données cliniques du patient et les règles de décisions de la base de connaissance et construit une critique de la prescription du médecin [Ebrahimi 2006].

## **2.2 Quelles connaissances dans les différents modèles de représentation des connaissances des GBPC ?**

Les GBPC peuvent être vus comme un ordonnancement de décisions et d'actions. Les formalismes de représentation développés visent à identifier ces éléments et à proposer des modalités d'ordonnancement dans le temps [De Clerc 2004]. On peut distinguer deux approches : dans la première, les recommandations sont des décisions dépendant des états du patient et des modalités de soins (*state-based modelling*) ; dans la deuxième, les recommandations sont des séquences d'actions et de décisions mises en œuvre pour atteindre un certain objectif (*plan-based modelling*).

L'approche EON [Musen 1996] par exemple propose une modélisation par les états du patients sous la forme de scénarii dans lesquels les actions et décisions vont entraîner un changement de l'état du patient au cours du temps. Le formalisme GLIF (*Guideline Interchange Format* [Ohno-Machado *et al.* 1998]), proposant une modélisation par la planification des soins, permet de modéliser les guides sous la forme de diagrammes algorithmiques en décrivant des étapes (*guideline step*) ordonnées dans le temps. Ces étapes sont décrites sous la forme de conditions, d'actions, de décisions et de modalités d'exécution des étapes (simultanéité, synchronisation).

### **2.3 Moyens mis en oeuvre pour aider à la structuration des guides**

La plupart de ces formalismes sont assez complexes et la formalisation manuelle des guides selon ces modèles est source de difficultés (compréhension du modèle) et de variabilité entre individus. Aussi pour supporter cette formalisation des connaissances dans les GBPC, des outils permettant le balisage du texte ont été créés, reposant sur des modèles documentaires (comme GEM — *Guideline Element Model* [shiffman 2000]).

La structuration, dans certains de ces éditeurs, est manuelle : l'utilisateur choisit les composants à structurer au moyen d'une interface graphique (GEM Cutter, Guide-X/Stepper, GMT/DELT/A, DeGeL/URUZ). Cette approche manuelle conduit à une grande variabilité dans la structuration, car l'interprétation des guides peut être différente d'un individu à l'autre.

D'autres outils proposent un traitement du document en préalable à la structuration, par repérage de marqueurs caractéristiques dans le texte (opérateurs déontiques dans G-DEE [Georg 2005]) ou par apprentissage [Kaiser 2007]. L'approche systématique présente l'intérêt de la reproductibilité. Cependant, les logiciels mis au point jusqu'ici restent limités : l'analyse ne va pas au-delà la phrase, alors que les conditions et les actions s'enchaînent de manière complexe et sont interdépendantes dans les textes.

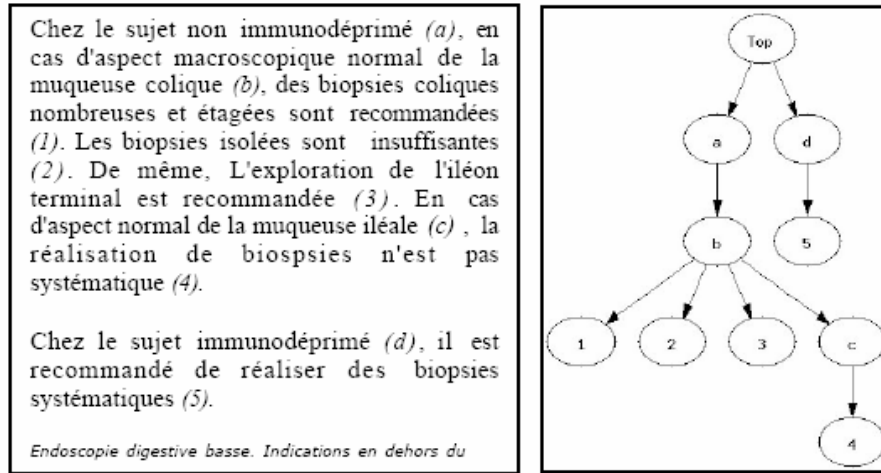
Le système GemFrame repose sur une analyse du texte pour présenter à l'utilisateur les conditions, les actions et les interactions entre ces différents éléments. Afin de pallier les limites des approches précédentes, le logiciel procède à une analyse du niveau discursif, ce qui permet de dépasser la limite de la phrase.

## **3 Modélisation des guides de bonnes pratiques cliniques : le système GemFrame**

Nous présentons ici le système GemFrame, dédié au repérage de séquences « conditions-recommandations » au sein des GBPC. L'originalité de ce système est de calculer la portée des segments conditionnels, ce qui constitue une réelle valeur ajoutée par rapport aux systèmes similaires limités à la phrase [Georg 2005].

### **3.1 Un Traitement en deux étapes**

Le traitement que nous proposons est constitué de deux étapes principales :  
1) repérage des segments élémentaires « condition » et « recommandations » ;  
2) calcul de la portée des segments conditionnels (ce qui revient à déterminer sous quelle condition une recommandation donnée est vraie). Au final, la structure générée automatiquement est un arbre XML dont les feuilles sont des segments « recommandations » et les noeuds des segments « conditionnels ». Tous les descendants d'un noeud condition *C* (c'est-à-dire une liste de recommandations ou de



**Figure 1 :** Du texte à l'arbre représentant la portée des segments conditionnels

conditions dans le cas de structures imbriquées) sont sous la portée de *c* (autrement dit, d'un point de vue véridictionnel, toutes les recommandations vraies uniquement sous cette condition et non de façon générale). La figure 1 montre un exemple d'arbre produit par le système.

L'ensemble du traitement est réalisé automatiquement. Le résultat doit néanmoins être validé de manière interactive par un expert : notre système constitue donc une « aide à la modélisation ». Dans ce qui suit, nous présentons les deux étapes de traitement et, de manière plus détaillée, les règles permettant de calculer la portée des segments conditionnels.

### 3.2 La segmentation élémentaire

La première étape consiste à découper le texte en segments élémentaires de type « condition » et « recommandation ». Nous repérons dans un premier temps les *marques linguistiques* exprimant la condition et la recommandation. Nous utilisons pour cela une base de règles qui exploitent des classes de *marqueurs linguistiques* (tels que des verbes comme *recommander*, *conseiller* ou des adjectifs tels que *nécessaire*, *important*, *etc.*) acquis semi-automatiquement à partir de régularités repérées en corpus.

Dans un second temps, il s'agit de délimiter les segments élémentaires. Afin de simplifier les traitements, nous avons fait le choix de ne pas segmenter en deçà du niveau de la phrase ou du niveau de la proposition, lorsque celle-ci est délimitée graphiquement<sup>4</sup>. C'est pourquoi cette segmentation s'appuie essentiellement sur des délimiteurs physiques. Un ensemble de règles étend les segments jusqu'à la présence de délimiteurs retenus comme pertinents (début paragraphe, fin de phrase, fin d'une énumération, *etc.*).

À l'issue de cette étape, les segments élémentaires conditions et actions ont été repérés. Le premier cadre de la figure 1 montre un exemple de ce qui est produit en sortie de cette étape (les segments sont numérotés).

### **3.3 Calcul de la portée des segments conditionnels.**

Une fois les segments élémentaires repérés, il faut repérer les cadres introduits par les segments « condition ». Cette étape repose en grande partie sur la position dans le texte des segments conditions repérés à l'étape précédente. Les positions respectives donnent lieu à une segmentation par défaut qui peut être contredite en présence de marques de rupture ou de continuité.

#### **3.3.1 Indices typo-dispositionnels pour le calcul des cadres**

Une étude manuelle de plusieurs guides [Bouffier, 2006] a montré que la position des segments condition joue un rôle déterminant pour la segmentation en cadres. Nous distinguons deux cas :

- L'introducteur est une expression non intégrée syntaxiquement (ce cas inclut les expressions détachées en début de phrase, mais aussi les titres et les amorces d'énumération).
- L'introducteur est une expression intégrée syntaxiquement à la phrase.

Les introducteurs en position détachée ont une propension à avoir une portée étendue (au-delà de la phrase courante) alors que les introducteurs en position intégrée ont une propension à avoir une portée minimale. Plus précisément, l'étude déjà citée a montré que la majorité des expressions détachées introduisent un cadre qui se ferme à la fin du paragraphe courant alors qu'en général, une expression intégrée à la phrase introduit un cadre qui se ferme à la fin de la phrase courante. Ceci permet de fonder l'analyse sur une segmentation par défaut.

1. Les titres ont une portée qui s'étend jusqu'au titre suivant de même niveau ;
2. Les amorces d'énumération ont une portée qui couvre tous les éléments de l'énumération ;
3. Dans le cas des expressions détachées : la segmentation par défaut est égale au paragraphe ;
4. Dans le cas des expressions intégrées : la segmentation par défaut est égale à la phrase.

Pour les cas 3 et 4, suivant les GBPC considérés, la segmentation par défaut couvre 50 à 80 % des cas. Elle est toutefois remise en cause quand la présence d'autres indices indique une continuité entre différents paragraphes ou, à l'inverse, une rupture au sein du paragraphe.

#### **3.3.2 Cas de remise en cause de la segmentation par défaut**

Nous présentons dans cette section, de manière non systématique, quelques cas où la segmentation par défaut est remise en cause du fait de la présence d'indices de rupture ou de continuité au sein du texte.

**Redondance entre le titre et le premier introducteur de condition (exception à la règle 3).** Nous avons observé des cas relativement fréquents où le titre est en partie redondant avec le premier introducteur. Bien que celui-ci soit en position détachée, sa portée dépasse le paragraphe et se confond avec celle du titre.

**Présence de marques de rupture discursive (exception à la règle 3).** Certaines marques discursives sont à même de signaler la fermeture prématurée d'un cadre avant la fin de paragraphe, alors même que ce cadre est introduit par une expression détachée. Ces marques peuvent marquer un contraste (à travers des formes lexicales comme *cependant*, *en revanche...*) ou une justification (*en effet*, *en fait...*)<sup>1</sup>.

**Présence de marques de relations anaphoriques.** Certaines relations anaphoriques sont des indices privilégiés pour signaler la continuation d'un cadre après la fin de la phrase, alors même que son introducteur est une expression détachée.

L'analyse de l'ensemble de ces éléments a été implémentée à travers un système multi-agent permettant de gérer les multiples contraintes intervenant dans le processus de structuration des GBPC. Le système ainsi constitué reçoit en entrée un GBPC et fournit en sortie un GBPC structuré d'après le modèle GEM (pour plus de détails concernant l'implémentation, voir [Bouffier 2007, Bouffier et Poibeau 2007]). Il est alors possible de comparer ce processus avec une structuration manuelle des GBPC, afin d'évaluer les performances du système.

## 4 Evaluation

Les outils aidant à la formalisation des GBPC doivent être capables de proposer une analyse des GBPC proche de celle d'un analyste humain. Pour appréhender les performances des outils, il convient de comparer leurs résultats à ceux d'un groupe d'experts sur des guides n'ayant pas servi lors du développement des outils.

### 4.1 Principes généraux sur l'évaluation

Plusieurs approches peuvent être envisagées pour l'évaluation. La première approche consiste à présenter à l'expert le texte balisé et à lui demander si ce balisage est correct. Le principal inconvénient de cette méthode est que le système propose déjà une interprétation du guide ; cela peut biaiser le jugement.

La seconde approche consiste à confronter l'annotation obtenue par l'outil à une référence obtenue suite à une annotation manuelle. Le travail d'annotation est fastidieux, surtout s'il est fait de manière croisée (chaque texte est annoté par plusieurs personnes et les conflits sont ensuite résolus). Malgré sa lourdeur, c'est cette approche qui donne la meilleure garantie quant à la validité des résultats.

Les GBPC étant souvent complexes et d'interprétation difficile, le recours à plusieurs annotateurs risque d'aboutir à plusieurs interprétations différentes. Il est

---

<sup>1</sup> Bien que ces marques soient généralement considérées comme des marques de cohésion, elles nous servent ici à identifier la fin du cadre, dans la mesure où la suite du texte correspond à une séquence étiquetée « Justification » suivant la DTD GEM.



possible de mesurer le niveau d'accord entre les experts (par le coefficient Kappa par exemple) mais cela n'indique pas quelle est l'interprétation juste et si l'outil s'en approche. Pour s'affranchir des variations dans l'interprétation des guides, il est possible de chercher le consensus : les désaccords entre annotateurs sont relevés, et on redemande aux annotateurs de réinterpréter le guide à la lumière des autres interprétations jusqu'à ce qu'un consensus apparaisse, pour produire une référence (*gold standard*).

La confrontation des résultats de l'outil au *gold standard* peut s'exprimer à l'aide des indicateurs de rappel (ratio d'information correcte extraite par rapport à l'information disponible) et de précision (ratio de l'information correcte extraite par rapport à la totalité de l'information extraite). Le système doit privilégier une mesure de rappel élevé (idéalement égale à 100%), si besoin au détriment de la précision, car il est plus facile d'enlever de l'information non pertinente que de réinterpréter la totalité d'un guide en raison d'un défaut de détection d'information pertinente.

## **4.2 Mise en oeuvre de l'évaluation de l'outil GemFrame**

Nous présentons dans cette section les résultats de l'évaluation.

### **4.2.1 Élaboration de la référence manuelle**

Un certain nombre de GBPC ont été annotés manuellement pour servir de base à l'étude et pour servir lors de l'évaluation. L'évaluation est faite sur des GBPC n'ayant pas servi lors de la mise au point du système. Globalement, chaque annotateur doit dériver à partir d'un GBPC un arbre XML conforme à la notation GEM. L'accord entre annotateurs est ensuite calculé en comparant le nombre de recommandations reliées à la condition ou à l'ensemble de conditions dont elle dépend.

L'accord entre annotateurs, calculé à partir d'un ensemble de 162 noeuds, est de 0,96 (157 segments correctement raccordés sur 162). Ce résultat montre la faisabilité de la tâche et une variation dans l'annotation réduite, au moins sur certains GBPC. Des expériences faites en comparant une annotation non-experte avec une annotation experte révèle seulement des différences minimales. Ceci permet de dégager deux conséquences : 1) l'annotation ne repose que légèrement sur des connaissances expertes et 2) elle peut être faite manuellement en s'appuyant sur les caractéristiques linguistiques et typo-dispositionnelles, dans la mesure où les connaissances du domaine ne sont que peu sollicitées.

D'autres études ont mis en évidence une plus grande variabilité entre annotateurs humains (cf. *supra*, introduction). Il serait intéressant de prolonger l'expérience pour voir dans quelle mesure ces résultats varient en fonction du GBPC analysé, des guides d'annotation fournis et des annotateurs concernés.

### **4.2.2 Résultats globaux**

On évalue le découpage en segments élémentaires en calculant des scores de rappel et de précision. On obtient alors les résultats suivants (la F-mesure est la moyenne harmonique du rappel et de la précision).

<b>GBPC : Cancer du sein</b>			
	<b>Conditions</b>	<b>Recommandations</b>	<b>Rattachements</b>
# présents	73	96	169
# retrouvés	60	88	
<i>rappel</i>	82,19	91,66	
# corrects	70	94	126
<i>précision</i>	95,89	97,91	74,55
<b>F-mesure</b>	<b>88,85</b>	<b>94,64</b>	

<b>GBPC : Critères chimio</b>			
	<b>Conditions</b>	<b>Recommandations</b>	<b>Rattachements</b>
# présents	70	107	177
# retrouvés	61	96	
<i>rappel</i>	87,14	89,71	
# corrects	65	104	136
<i>précision</i>	92,85	97,19	76,83
<b>F-mesure</b>	<b>90,05</b>	<b>93,29</b>	

<b>GBPC : Dénutrition des personnes âgées</b>			
	<b>Conditions</b>	<b>Recommandations</b>	<b>Rattachements</b>
# présents	75	107	182
# retrouvés	62	100	
<i>rappel</i>	82,66	93,45	
# corrects	73	106	162
<i>précision</i>	97,33	99,06	89,01
<b>F-mesure</b>	<b>89,34</b>	<b>93,29</b>	

<b>GBPC : AOMI</b>			
	<b>Conditions</b>	<b>Recommandations</b>	<b>Rattachements</b>
# présents	60	91	151
# retrouvés	45	65	
<i>rappel</i>	75	71,42	
# corrects	59	88	107
<i>précision</i>	98,33	96,70	70,86
<b>F-mesure</b>	<b>85,08</b>	<b>82,14</b>	

Pour les rattachements des recommandations aux conditions (dernière colonne), comme on compare le nombre d'éléments correctement rattachés par rapport à la référence, on obtient juste un score de précision (*accuracy*).

### 4.2.3 Évaluation du découpage en segments élémentaires

Les résultats sont relativement élevés, aussi bien pour la reconnaissance des recommandations que pour la reconnaissance des conditions. L'apport des connaissances du domaine n'est pas évident à la vue des résultats. Cette information est cependant pertinente quand il s'agit d'étiqueter des éléments isolés, par exemple des titres correspondant à des pathologies. Par exemple le titre hypertension artérielle est équivalent à une condition introduite par *en cas de*. Il est donc important de reconnaître ces cas et de les analyser correctement dans la mesure où plusieurs recommandations sont susceptibles d'être sous la portée de cette condition. Cette analyse ne peut être faite sans connaissances du domaine.

Le nombre et la nature des titres diffèrent profondément d'un guide à l'autre. Quand le nombre de titres équivalent à une instruction conditionnelle est important, l'impact des connaissances du domaine n'est pas négligeable dans la mesure où la mauvaise analyse des titres peut faire manquer l'analyse de toute une série de recommandations. Enfin, on peut constater que tous les titres n'ont pas la même importance (certains sont des noeuds centraux, ils contrôlent davantage de recommandations, *etc.*) mais il est difficile de tenir compte de cet aspect dans l'évaluation.

### 4.2.4 Évaluation de la portée des conditions

La portée des conditions est analysée correctement dans 77 % des cas (ce chiffre est obtenu en extrayant, à partir de l'annotation manuelle, l'ensemble des couples simples « condition/recommandation ». On compare ensuite le nombre de couples communs entre la référence obtenue par l'étiquetage manuel et les données produites par le système. On obtient ainsi le taux de précision de l'analyseur).

Ce résultat est encourageant, surtout si on prend en compte l'ensemble des paramètres impliqués lors de l'analyse du discours. Dans la plupart des cas, la portée de la condition est obtenue à partir de l'application des règles par défaut. Cependant, quelques cas importants sont résolus grâce aux règles fondées sur l'analyse des exceptions, qui permettent d'aller en deçà ou au-delà de la portée par défaut.

Le système échoue particulièrement en cas de portée étendue, quand celle-ci est exprimée par des marques de cohésion liées au vocabulaire du domaine (utilisation de synonymes, d'hyponymes ou d'hyperonyme) ou de structures complexes (anaphores nominales, structures syntaxiques complexes). Résoudre ces cas rares reviendrait à introduire de nombreuses connaissances du domaine dans le système, avec tous les risques que cela comporte. Nous avons volontairement écarté cette voie dans la mesure où l'on souhaite garder un système relativement portable, faisant un usage minimal des connaissances du domaine.

### 4.2.5 Utilisabilité de la solution

Au-delà de l'évaluation qualitative, il serait intéressant de compléter l'approche avec des tests d'utilisabilité. On peut d'ores et déjà noter plusieurs points qui sont primordiaux en situation de consultation : l'outil est le seul à proposer une analyse allant au-delà de la phrase ; il peut vraiment rendre compte des ensembles de

conditions que le médecin doit prendre en compte. Les informations sont structurées sous forme d'arbre, ce qui correspond à un mode de raisonnement pertinent pour les professionnels de santé. L'accès à l'information peut alors se faire à travers un modèle hypertexte comme dans OncoDoc [Bouaud *et al.* 1999] ou de manière multimodale, afin de permettre plusieurs modes d'accès complémentaires [Mondary *et al.* 2007].

## 5 Conclusion

Nous avons présenté dans cet article un outil d'aide à la modélisation des GBPC. Ces documents nécessitent une modélisation, mais ce travail manuel est long, fastidieux et variable d'un individu à l'autre. L'automatisation de la tâche permet de s'affranchir partiellement de la variation individuelle en proposant une méthode fixe, définie et reproductible. Les résultats obtenus sont satisfaisants eu égard à la difficulté de la tâche. Une étude plus précise des cas d'échecs serait nécessaire afin d'affiner l'outil. Il faut toutefois noter que certains cas semblent difficilement automatisables; surtout si l'on veut garder une certaine généricité dans les traitements. L'outil reste donc une aide à la modélisation et il ne prétend en aucun cas remplacer l'analyste humain. Il doit au contraire permettre à celui-ci de se concentrer sur les cas difficiles en automatisant le traitement des cas les cas simples.

## Remerciements

Les recherches présentées dans cet article sont partiellement financées par le programme pluri-formations ActMed (à l'Université Paris 13, entre les laboratoires LDI, LIM&Bio et LIPN). Les travaux du LIPN s'inscrivent également dans le cadre du projet ANR TextCoop, dont les partenaires sont (outre le LIPN), l'IRIT et la société Sinequa.

## Références

- BOUAUD J, SEROUSSI B, ANTOINE E.C. ONCODOC : une approche documentaire de l'aide à la décision. Document numérique. 1999, vol. 3, n° 3-4: pp. 61-79.
- BOUFFIER A.. From Texts to Structured Documents: The case of Health Practice Guidelines. *International Semantic Web Conference, Doctoral Consortium*, Busan, Corée, 2007.
- Bouffier A., POIBEAU T. Automatic Structuring of Practice Guidelines Using the GEM DTD. *BioNLP Workshop, Association for Computational Linguistics*, Prague, République Tchèque, 2007.
- BOXWALA AA, PELEG M, TU S, OGUNYEMI O, ZENG QT, WANG D, PATEL VL, GREENES RA, SHORTLIFFE EH. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. *J. Biomed. Inform.*. 2004 Jun; 37(3): 147-61.
- CABANA MD, RAND CS, POWE NR, WU AW, WILSON MH, ABBOUD PA, RUBIN HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999 Oct 20; 282(15):1458-65.
- CODISH S, SHIFFMAN RN. A model of ambiguity and vagueness in clinical practice guideline

- recommendations. *AMIA Annu. Symp. Proc.* 2005;:146-50
- DE CLERCQ PA, BLOM JA, KORSTEN HH, HASMAN A. Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artif. Intell. Med.* 2004; 31(1):1-27.
- ELKIN PL, PELEG M, LACSON R, BERNSTAM E, TU S, BOXWALA A, GREENES R, SHORTLIFFE EH. Toward the standardization of electronic guidelines. *MD Comput.* 2000 Nov-Dec;17(6):39-44.
- EBRAHIMINIA V, RIOU C, SEROUSSI B, BOUAUD J, DUBOIS S, FALCOFF H, VENOT A. Design of a decision support system for chronic diseases coupling generic therapeutic algorithms with guideline-based specific rules. *Stud Health Technol Inform.* 2006;124:483-8.
- GEORG G, COLOMBET I, JAULENT MC. Structuring Clinical Guidelines through the Recognition of Deontic Operators. *Stud Health Technol Inform.* 2005; 116:151-6.
- KAISER K, AKKAYA C, MIKSCH S. How can information extraction ease formalizing treatment processes in clinical practice guidelines? A method and its evaluation. *Artif Intell Med.* 2007 Feb; 39(2): 151-63.
- MONDARY T., BOUFFIER A., NAZARENKO A. Between browsing and search, A new model for navigating through large documents. *EuroCognitive Sciences*, Delphes, Grèce, 2007.
- MUSEN, M.A., TU, S.W., DAS, A.K., AND SHAHAR, Y. EON: A component-based approach to automation of protocol-directed therapy. *J. of the AMIA.*3(6): 367-388, 1996.
- OHNO-MACHADO L, GENNARI JH, MURPHY SN. The guideline interchange format: a model for representing guidelines. *J. of the AMIA.* 5(4):357-372, 1998.
- PATEL VL, AROCHA JF, DIERMEIER M, GREENES RA, SHORTLIFFE EHJ. Methods of cognitive analysis to support the design and evaluation of biomedical systems: the case of clinical practice guidelines. *Biomed. Inform.* 2001 Feb; 34(1): 52-66
- SÉROUSSI B, BOUAUD J, DRÉAU H, FALCOFF H, RIOU C, JOUBERT M, SIMON C., VENOT A. ASTI: a guideline-based drug-ordering system for primary care. *Medinfo.* 2001; 10(Pt 1): 528-32.
- SHAHAR Y, YOUNG O, SHALOM E, GALPERIN M, MAYAFFIT A, MOSKOVITCH R, HESSING A. A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. *J. Biomed. Inform.* 2004 Oct; 37(5): 325-44.
- SHALOM E, SHAHAR Y. A graphical framework for specification of clinical guidelines at multiple representation levels. *AMIA Annu. Symp. Proc.* 2005;: 679-83.
- SHIFFMAN RN, GREENES RA. Improving clinical guidelines with logic and decision-table techniques: application to hepatitis immunization recommendations. *Med. Decis. Making.* 1994 Jul-Sep; 14(3): 245-54.
- SHIFFMAN RN. Representation of clinical practice guidelines in conventional and augmented decision tables. *J. Am. Med. Inform. Assoc.* 1997 Sep-Oct; 4(5): 382-93
- SHIFFMAN RN, BRANDT CA, LIAW Y, CORB GJ. A design model for computer-based guideline implementation based on information management services. *J. Am. Med. Inform. Assoc.* 1999 Mar-Apr; 6(2): 99-103
- SHIFFMAN RN, KARRAS BT, AGRAWAL A, CHEN R, MARENCO L, NATH S. GEM: a proposal for a more comprehensive guideline document model using XML. *J. Am. Med. Inform. Assoc.* 2000 Sep-Oct; 7(5): 488-98.
- SHIFFMAN RN, SHEKELLE P, OVERHAGE JM, SLUTSKY J, GRIMSHAW J, DESHPANDE AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. *Ann. Intern. Med.* 2003 Sep 16; 139(6): 493-8
- SHIFFMAN RN, MICHEL G, ESSAIHI A, THORNQUIST E. Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation. *J. Am. Med. Inform. Assoc.* 2004 Sep-Oct; 11(5): 418-26