

Annotation automatique de relations de contrôle dans des spécifications des besoins informatiques

Jorge García-Flores, Elena Ivanova,
Jean-Pierre Desclés, Brahim Djoua

Université de Paris-Sorbonne – LaLICC
{jorge.gflores ; jpdescles ; elena.ivanova ; bdjioua}@paris4.sorbonne.fr

Résumé

La conception de logiciels est un processus technologique complexe, qui nécessite d'être assisté par des outils de traitement automatique des langues. Cet article présente une méthode pour l'annotation de relations discursives de contrôle dans des textes de spécification de besoins informatiques (SBI). La méthode vise à distinguer les actions contrôlées par le système de celles contrôlées par son environnement, ce qui permet d'établir de façon claire les limites et les responsabilités d'un système informatique. Notre méthode fait appel à la sémantique discursive pour analyser les moyens d'expression du contrôle dans un corpus de SBI industrielles ; l'expression du contrôle est identifiable par la présence, dans un certain contexte, de marqueurs linguistiques exprimés par des règles dites d'Exploration Contextuelle. La dernière partie montre le processus d'annotation automatique de la notion de contrôle par le système EXCOM et termine par la présentation d'un début d'évaluation de cette méthodologie.

Mots-clés : relations de contrôle, spécifications des besoins, extraction d'information, contrôle, sémantique, annotation automatique.

Abstract

This paper presents a technique for the automatic extraction of control sentences from software requirements specifications (SRS). Our aim is to annotate action sentences from an SRS document, and to recognize if those actions are machine controlled or user controlled, in order to better understand the limits and responsibilities of a software system and to support the Requirements Engineering process. We present a linguistic analysis of control markers and a technique to automatically annotate control sentences by means of contextual exploration rules. These rules are implemented in the contextual exploration engine Excom and applied to a set of commercial SRS.

Keywords : semantic filtering, natural language requirements, requirements engineering, control.

1. Introduction

La conception de logiciels est un champ d'application possible pour les outils de traitement automatique des langues. Dans le passé, des outils d'extraction d'information ont été utilisés pour supporter la modélisation de systèmes informatiques. Étant donné que la spécification des besoins informatiques (SBI) se réalise, dans la majorité des cas, par la rédaction de textes (cahier de charges, spécification de besoins, cas d'utilisation, etc.), il nous semble donc nécessaire de traiter ces documents avec des analyses linguistiques sémantiques pour pouvoir repérer des entités, des relations discursives et des ambiguïtés utiles à la conception d'un système informatique. On distingue trois types d'applications TAL dans le champ du génie logiciel, et plus précisément dans l'Ingénierie de Besoins (García Flores, 2004) : 1) aide à l'élicitation de besoins via l'extraction d'information à partir de comptes rendus, entretiens

avec l'utilisateur ou verbalisations ergonomiques; 2) aide à la validation (syntaxique ou sémantique) de besoins, et 3) l'extraction des données pour capitaliser des connaissances utiles à la modélisation du système. Ces trois types d'applications ont un souci en commun qui est de profiter de la richesse sémantique du texte pour assister un processus technologique complexe comme la conception d'un logiciel. Pour nous, cette sémantique est de nature discursive et se réalise par les contenus des textes que sont les spécifications des besoins.

La notion de contrôle joue un rôle essentiel dans la démarche d'Ingénierie de Besoins (Zave *et al.*, 1997). En effet, spécifier les besoins consiste à décrire soit des actions que le système effectue sur son environnement, soit des actions venant de cet environnement et devant lesquelles le système doit agir. Nous appelons environnement le contexte social, organisationnel et technologique dans lequel le système vient s'insérer. Distinguer les actions contrôlées par le système de celles contrôlées par l'environnement est nécessaire pour établir de façon claire les limites et les responsabilités d'un système informatique. Plus précisément, ce travail a pour but de répondre automatiquement, à l'aide de la sémantique discursive, à la question *Qui Contrôle Quoi ?* Par exemple :

- 1) *L'utilisateur peut composer une nouvelle requête à l'aide du langage de requête fourni par le SGBD. Il l'enregistre et la requête est ajoutée dans le référentiel DPST à la liste des autres requêtes.*

Tout d'abord, notre but est de repérer les verbes qui expriment effectivement des actions. Ensuite, il faut déterminer l'agent contrôleur et l'entité affectée par ces actions. Enfin, il faut déterminer si la responsabilité des actions appartient au système ou à son environnement. Le tableau 1 présente les informations que nous cherchons à repérer de façon automatique à partir de cet exemple.

<i>Action</i>	<i>Contrôleur</i>	<i>Entité affectée</i>	<i>Instrument</i>	<i>Domaine de contrôle</i>
composer	utilisateur	requête	langage de requête	environnement
enregistrer	Il -> l'utilisateur	l' -> la requête		environnement
ajouter	∅ -> le système	requête		système

Tableau 1. Repérage du contrôle dans l'exemple 1

Notre démarche s'appuie sur une analyse linguistique des moyens d'expression du contrôle à partir d'un corpus de textes de SBI industrielles qui spécifient des systèmes conçus dans des institutions comme l'Union Européenne, La Poste, Peugeot-Citroën, et Renault. Plusieurs systèmes d'annotation des textes de SBI existent dans la littérature ; nous citons notamment ceux pour l'aide à la rédaction (Ben-Achour, 1999), la désambiguïsation lexicale et syntaxique (Fantechi *et al.*, 2003), l'extraction de concepts (Sawyer *et al.*, 2004) ou la validation sémantique de besoins (Gervashi *et al.*, 2002). La méthodologie d'annotation linguistique automatique utilisée dans cet article pour le repérage des notions d'action, de contrôleur, d'entité affectée est celle dite d'Exploration Contextuelle (Desclés, 1991).

Dans la suite de l'article, nous présentons dans un premier temps la problématique et l'introduction du cadre théorique ; puis vient l'analyse et l'interprétation des observables linguistiques. Nous poursuivons par la présentation de la méthode d'Exploration Contextuelle à travers ses règles linguistiques pour le repérage de la notion contrôle pour finir par la présentation informatique du système EXCOM ainsi que les premiers résultats et un début d'évaluation. Une conclusion et des perspectives futures viennent clore cet article.

2. La notion de contrôle

2.1. Relations de contrôle dans les textes de SBI

Le but de ce travail est d'extraire de façon automatique les relations de contrôle dans un texte de SBI. Nous appelons relation de contrôle celle où un agent a la capacité de déclencher, interrompre ou arrêter une action qui peut affecter un patient, éventuellement à l'aide d'un instrument (Desclés *et al.*, 1991). Dans un texte de SBI, il y a deux types d'entités capables de contrôle : celles qui font partie de l'environnement du système (par exemple, *l'utilisateur*), et celles qui représentent le système, lui-même étant une entité capable de contrôle (par exemple, *la pré-installation*). Prenons un exemple extrait de notre corpus :

- 2) *La pré-installation porte sur les stations touchées par la planification et pour ces stations, prend en compte les versions de profil planifiées par la planification automatique. La pré-installation ne se déclenche automatiquement que si le contrôle de validité ne retourne pas d'erreur. L'utilisateur peut visualiser le compte-rendu de la planification automatique et l'éditer.*

La valeur de contrôle d'une action est déterminée par la capacité de contrôle de son agent plus la « propension » au contrôle du verbe. Dans le domaine de l'exemple ci-dessus, une *pré-installation* a une certaine capacité de contrôle puisqu'elle peut *se déclencher automatiquement*. Par contre, la signification du verbe *porte* annule cette capacité quand elle *porte sur les stations*. La capacité de contrôle d'une entité étant généralement relevant du domaine extralinguistique, le traitement visé doit recevoir en entrée une liste d'entités typées selon leur capacité de contrôle et leur appartenance au domaine du système ou à celui de l'environnement.

2.2. La notion de contrôle dans la GAC

Dans le modèle de la Grammaire Applicative et Cognitive (GAC), le contrôle est défini comme la capacité de déclencher et d'interrompre une activité ou une action. Parmi les niveaux de représentation dans la GAC (Desclés, 1990), le niveau cognitif contient des représentations sémantico-cognitives (des schèmes) qui représentent des significations des verbes et s'organisent à partir des primitives sémantico-cognitives. L'introduction des primitives s'appuie sur l'hypothèse que les catégories grammaticales et lexicales des langues sont ancrées sur des catégorisations opérées par les processus cognitifs de perception, action sur l'environnement et par prise en compte d'une certaine intentionnalité. L'hypothèse clarifie trois types de situations qui peuvent être désignées par un verbe :

- Situations statiques représentent des états sans changement, ni mouvement.
- Situations cinématiques représentent des actions, des processus, autrement dit des changements ou des mouvements qui transforment une situation statique initiale à une situation statique finale.
- Situations dynamiques représentent les situations cinématiques contrôlées et/ou effectuées par un agent contrôleur avec un certain degré d'intentionnalité qui peut viser la situation finale comme un but.

Aux trois types de situations correspondent trois groupes de primitives: statiques, cinématiques et dynamiques. Parmi les primitives dynamiques, la primitive CONTR est introduite pour tenir compte de l'intentionnalité et exprimer la capacité de contrôler une action. Nous tenons particulièrement à distinguer les situations statiques des situations évolutives (c'est-à-dire, cinématiques et dynamiques). Cette distinction est fondamentale dans

l'organisation des connaissances linguistiques qui nous permettront de repérer automatiquement le contrôle.

3. Analyse linguistique de contrôle dans les textes de SBI

Le texte de SBI est un texte hautement structuré, dont la rédaction est généralement guidée par une grille. Il n'est pas rare que la taille d'un tel document atteigne plusieurs centaines de pages. Le corpus utilisé dans le cadre de notre travail se compose de douze textes qui spécifient les besoins pour construire des systèmes informatiques industriels (Peugot-Citroën, La Poste, système de radio européen, etc.). La taille d'un texte varie de 20 à 150 pages.

L'analyse linguistique de contrôle se base principalement sur les verbes. L'hypothèse fondamentale consiste à voir le contrôle d'une situation décrite dans une phrase, comme un résultat d'interaction de la propension sémantique au contrôle du verbe (encodée dans son schème sémantico-cognitif), de la capacité de contrôle de l'agent et des compléments circonstanciels capables d'ajouter ou d'enlever le contrôle malgré les deux premières conditions. Notre analyse des moyens d'expression du contrôle dans le corpus reprend des concepts propres à la GAC et les adapte à la problématique particulière du repérage du contrôle dans un texte de SBI. Par exemple, le fait que le système informatique devienne une entité capable de contrôle illustre bien une telle adaptation.

3.1. Verbes évolutifs

Prenons les phrases dont le prédicat est un verbe conjugué. Certains verbes ont de manière intrinsèque un facteur discriminant pour que l'on puisse juger s'ils requièrent ou non un agent contrôleur (Soh, 2002). L'opposition entre situations statiques et situations évolutives nous permet de reconnaître les verbes à la voix active qui renvoient à des actions, et de les distinguer de ceux qui renvoient à des états.

- 3) *L'utilisateur choisit les versions de livrables qu'il désire intégrer dans la version de profil.*
- 4) *Ce fichier de données structurées comporte des références à des objets multimédia.*

Dans l'exemple 3, le verbe *choisir* désigne une situation évolutive; il possède une forte propension au contrôle : il attend généralement un agent contrôleur. L'analyse des textes de SBI nous permet de classer des verbes comme *ajouter*, *sélectionner*, *modifier*, *signaler*, *intégrer*, etc. comme des verbes évolutifs. Par contre, l'exemple 4 décrit une situation statique avec une relation d'appartenance des références au fichier de données. Des verbes comme *contenir*, *exister*, *avoir*, *consister*, etc. font partie d'une liste de verbes statiques. Dans le repérage du contrôle, nous ne nous intéressons qu'aux verbes évolutifs.

3.2. Constructions passives

Parmi les nombreuses constructions passives qu'on trouve dans les textes de SBI, nous nous intéressons particulièrement au passif processuel (Desclés, 1994) dans la mesure où il désigne un processus en train de se dérouler (voir l'exemple 5). Il est important de distinguer le passif processuel des constructions passives qui expriment une propriété résultative, comme celle de l'exemple 6 :

- 5) *Il l'enregistre et la requête est ajoutée à la liste des autres requêtes.*

- 6) *L'optimisation principale est liée à l'utilisation de la mémoire RAM.*

Nous distinguons ces deux constructions passives par la sémantique du verbe. Nous regroupons des verbes comme *lier, baser, destiner, fonder*, etc. dans une classe de verbes qui, au passif, renvoient à des situations statiques et donc ne doivent pas être prises en compte pour le repérage du contrôle. Dans les constructions passives, la place du sujet syntaxique est occupée par une entité affectée. Le contrôleur est également présent dans le passif agentif comme un complément d'agent introduit par la préposition *par*. Cependant, la plupart de constructions passives de notre corpus sont non agentives, l'agent étant effacé (l'exemple 5).

Pour différencier, dans le corpus, les actions contrôlées par le système de celles contrôlées par l'environnement, nous avons recours à l'information extra linguistique. Ainsi les verbes *ajouter, configurer, exécuter, remplacer, informer*, etc. au passif non agentif sont mis en relation avec l'entité « le système » comme l'agent contrôleur, tandis que les verbes *choisir, consulter, sélectionner, saisir, trier, valider*, etc. utilisés à la même forme du passif non agentif sont eux mis en relation avec l'« agent l'environnement ». Cette règle heuristique est particulière au contexte de l'Ingénierie des Besoins, où un système informatique est vu comme une entité capable de contrôle et d'exécution d'actions distinctes de celles effectuées par un utilisateur.

3.3. Verbes modaux

Quand l'agent contrôleur est effacé, la réponse à la question *Qui Contrôle Quoi ?* nécessite d'autres marqueurs linguistiques. Le verbe modal d'obligation (*devoir*) est utilisé pour exprimer une situation sous contrôle du système (l'exemple 7), tandis que les verbes de volonté (*vouloir, désirer, souhaiter*) précèdent les verbes qui renvoient à des actions contrôlées par l'environnement (l'exemple 8). Cette opposition des verbes modaux est propre aux textes de SBI, où le système représente un moyen (instrument) qui permet aux utilisateurs de toucher leurs buts (d'où les notions de l'obligation et de la volonté).

- 7) *L'identité fonctionnelle de l'appelant doit être affichée lorsqu'elle est disponible.*
 8) *L'utilisateur désire valider une diffusion pour la rendre diffusable et donc non modifiable.*

Un autre indice de contrôle est la construction *permettre de + infinitif*. Dans l'exemple 9, le sujet syntaxique *une boîte de dialogue* désigne un instrument à l'aide duquel l'utilisateur (dont l'anaphore est *lui*¹) pourrait effectuer l'action exprimée par l'infinitif. Notons que cette construction représente toujours une action contrôlée par l'environnement et le système y joue le rôle d'instrument.

- 9) *Une boîte de dialogue lui permet de saisir un message explicatif.*

3.4. Moyens non prédicatifs

Dans la partie introductive de chaque document SBI de notre corpus de travail sont définis une liste d'entités capables de contrôle. Cette liste comporte des noms désignant l'utilisateur (*utilisateur, consommateur*, etc.) et des noms désignant le système (*logiciel, système, outil*). Pour qu'une phrase soit annotée comme situation contrôlée, une entrée de cette liste doit être soit le sujet de la phrase active, soit le complément d'agent s'il s'agit du passif agentif. La

¹ Le problème d'anaphore n'a pas été étudié dans le présent travail.

seule exception trouvée concerne le passif non agentif. Dans le cas d'un passif non agentif processuel la phrase est annotée comme situation contrôlée.

Des constructions adverbiales comme *automatiquement*, *de façon automatique*, etc. à cause de leur sémantique permettent de repérer des situations contrôlées par le système.

4. Règles d'exploration contextuelle

La stratégie du TAL que nous utilisons afin de repérer automatiquement le contrôle est l'exploration contextuelle, proposée par (Desclés, 1991) pour la désambiguïsation lexicale, et utilisée ensuite pour des tâches comme le repérage des relations entre concepts (LePriol, 2000) ou des relations de causalité (Jackiewicz, 1998). L'exploration contextuelle (EC) s'appuie sur un raisonnement abductif qui consiste à poser une hypothèse et à évaluer sa plausibilité en cherchant les indices observables. L'EC organise les marqueurs linguistiques sous forme de règles déclenchées par la présence d'un indicateur (hypothèse abductive).

Règle	Indicateur	←Indices→ ²	Exemple
R1	Verbe conjugué non statique	←entité contrôleur	<i>L'utilisateur choisit les versions de livrables.</i>
R2	Infinitif non statique	←entité contrôleur ←verbe modal capacité	<i>L'administrateur peut ensuite modifier cette composition.</i>
R3	Participe passé non statique	←être (conjugué) ¬ ← si ¬ ← se	<i>Une vérification est faite dans DEMETER pour s'assurer que le projet n'existe pas déjà.</i>
R4	Participe passé non statique	←être (conjugué) par→ entité contrôleur → ¬ ← si ¬ ← se	<i>Les deux plaques sont serrées par une vis qui bride le vitre sur le curseur en X, Y.</i>
R5	Participe passé non statique	←devoir (conjugué) ←être (infinitif) ¬ ← si ¬ ← se	<i>Une connexion en phonie doit être établie immédiatement après la tonalité d'avertissement.</i>
R6	Participe passé non statique	←devoir pouvoir (conjugué) ←être (infinitif) par→ entité contrôleur → ¬ ← si ¬ ← se	<i>Cette opinion peut être également définie par la DPST.</i>
R7	Verbe modal telique conjugué		<i>L'utilisateur désire valider une diffusion pour la rendre diffusible.</i>
R8	<i>permettre</i> (conjugué)	Infinitif non statique→ de→ [←lui leur] [à au aux→ entité contrôleur →]	<i>Une boîte de dialogue lui permet de saisir un message explicatif.</i>
R9	<i>automatiquement</i>	←se	<i>La pré-installation se déclenche automatiquement.</i>

Tableau 2. Récapitulatif des règles d'EC pour le repérage automatique de contrôle

² Les symboles ← et → représentent l'espace de recherche: contexte gauche jusqu'au début de la phrase, contexte droite jusqu'à fin de la phrase respectivement, ¬ représente la négation exprimée ici par l'absence des indices et | représente l'opération d'alternative exprimée ici par l'union des classes d'indices.

La règle cherche des indices de plausibilité dans le contexte de l'indicateur. Si l'hypothèse est confirmée, la règle attribue une valeur sémantique sous forme d'annotation. Dans (García Flores, 2004), nous avons analysé l'EC à la lumière des autres stratégies du TAL utilisées pour supporter la démarche d'IB. L'application de ces règles dans ces textes implique son instanciation dans l'un des systèmes d'EC développés au sein de l'équipe LaLICC. Une règle d'EC fait appel à la fois à des indicateurs et des indices, répertoriés dans des classes de marqueurs. Comme le contrôle est un composant primitif d'un prédicat linguistique, les indicateurs déclencheurs des règles de contrôle seront des verbes (voir tableau 2).

Nous distinguons trois groupes de règles. Le premier groupe (règles R1 et R2) repère les traces de contrôle exprimées par des verbes évolutifs à la voix active ayant comme sujet (donc dans le contexte gauche) une entité capable de contrôle. Le deuxième groupe (règles R3 à R6) se déclenche par des verbes évolutifs au passif. Comme cela a été dit dans la section précédente, la voix passive change la propension au contrôle de certains verbes : certains verbes deviennent statiques. R3 et R4 reconnaissent respectivement le passif non agentif et le passif agentif sans modalité, tandis que R5 et R6 cherchent un passif progressif précédé par un verbe modal. L'indice $\neg \leftarrow si$ interdit la présence de la conjonction *si* dans le contexte gauche afin d'éviter la conditionnelle. L'indice $\neg \leftarrow se$ permet de distinguer le passif du passé composé d'un verbe réfléchi et de ne pas traiter le dernier. Le troisième groupe (règles R7 à R9) sert à repérer la modalité de volonté, l'adverbe *automatiquement* et la construction *permettre de*.

5. Annotation automatique des situations de contrôle avec Excom

5.1. Processus d'annotation

Pour l'annotation des textes de SBI de notre corpus, nous utilisons le système d'annotation par Exploration Contextuelle EXCOM (Djioua *et al.*, 2006). L'annotation (voir figure 1) procède en quatre grandes étapes : d'abord une opération de prétraitement est déclenchée, où chaque texte est converti au format Unicode UTF-8 ainsi qu'une opération de nettoyage des caractères interdits en XML. Ensuite est déclenchée une opération de segmentation des textes nettoyés en sections, paragraphes, et phrases selon des critères typographiques et structuraux, effectuée à l'aide de l'outil de segmentation *SegaTex* (Mourad, 2001).

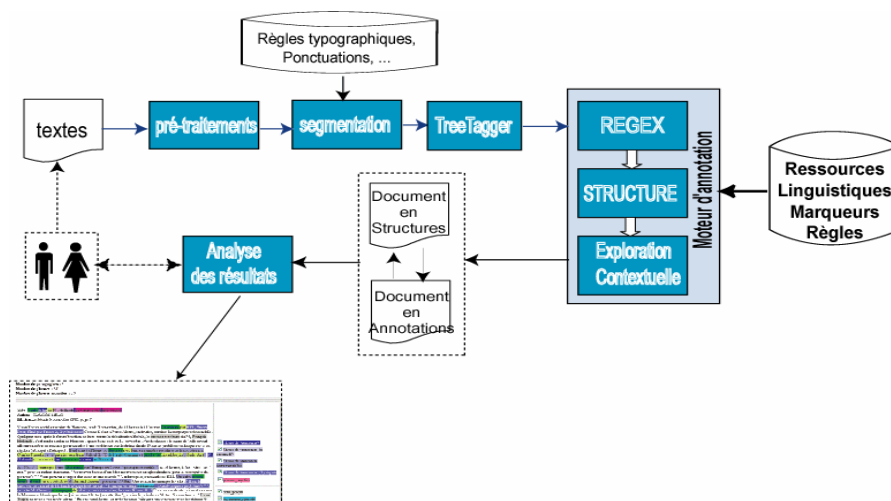


Figure 1. Architecture du processus d'annotation de situations de contrôle

Une fois le texte segmenté, les verbes évolutifs sont pré-annotés à l'aide de TreeTagger (Schmidt, 1994), et on ne retiendra que les verbes évolutifs (voir section 5). Dans la quatrième étape, EXCOM applique les règles d'EC pour le repérage du contrôle sur le texte afin d'y annoter les segments qui désignent des situations de contrôle. Il faut préciser que, grâce à une spécificité des textes de SBI, où la hiérarchie d'utilisateurs est toujours spécifiée au début du document, le processus d'annotation reçoit en entrée une liste d'entités capables de contrôle et, en conséquence, n'a pas besoin de module de reconnaissance d'entités nommées.

Les étapes qui posent le plus de problèmes au processus d'annotation sont la segmentation et l'annotation des verbes évolutifs. La segmentation est particulièrement délicate parce que les SBI sont des textes hautement structurés, avec de nombreux sous-titres, tableaux et énumérations, ce qui peut entraîner des erreurs dans un bon découpage des phrases. Quant à l'annotation des verbes évolutifs, nous avons deux possibilités : 1) soit de prélever une liste de verbes évolutifs et l'ajouter aux bases de marqueurs linguistiques comme on fait généralement avec la méthode d'EC, 2) soit d'utiliser un analyseur morphosyntaxique (dans ce cas, TreeTagger) pour le repérage des verbes. Nous avons choisi la deuxième option. Ce choix est justifié par la spécificité de la notion de contrôle par rapport à d'autres tâches d'EC. Comme il a été montré précédemment (voir section 2.1), tout verbe qui ne désigne pas clairement une situation statique est, potentiellement, un marqueur de contrôle. Il nous a paru plus économique de *souçonner* tout verbe n'appartenant pas à la classe de verbes statiques comme étant un verbe de contrôle, que de construire de longues listes avec tous les verbes de la langue, sauf ceux considérés comme statiques. Ce même principe d'économie faciliterait l'annotation des SBI dans des langues autres que le français, puisque les listes de marqueurs sont réduites. Par contre, l'inconvénient le plus important est l'utilisation de TreeTagger qui consiste au fait que toute erreur issue de l'annotation des verbes affecte la qualité du reste du processus. Le résultat d'une annotation automatique de la notion de contrôle se présente comme :

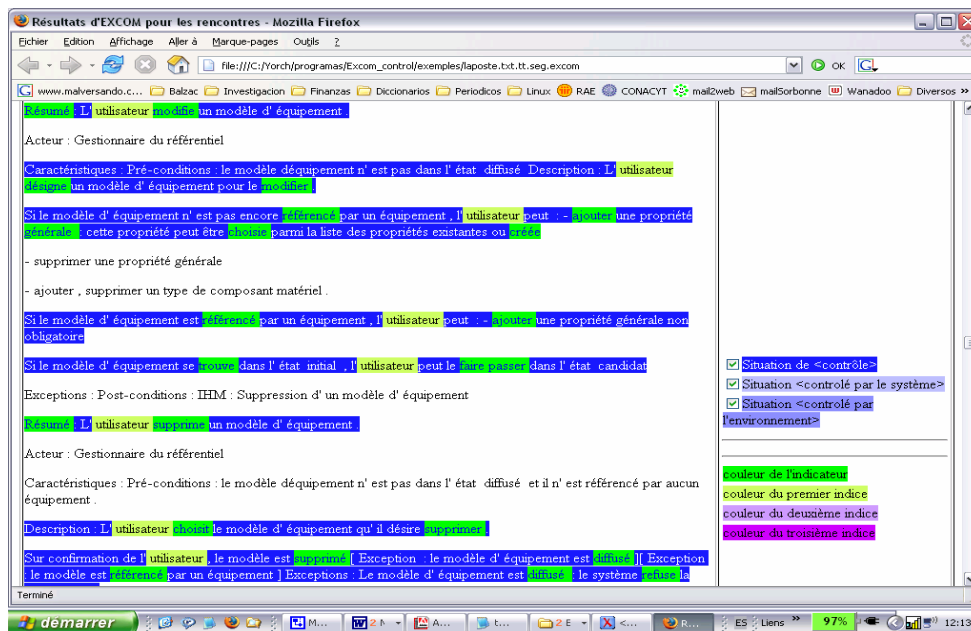


Figure 2. Affichage d'un texte de SBI annoté par EXCOM

5.2. Évaluation préliminaire

Une évaluation préliminaire des trois premières règles pour le repérage des relations de contrôle a été effectuée sur un texte de SBI. Le document de test spécifie les besoins d'un système de distribution de logiciels pour La Poste. Le document comporte 150 pages et il a été préalablement annoté à la main. La sortie d'EXCOM a été comparée aux annotations manuelles supposées être sans erreur. Parmi les 1300 phrases du document, EXCOM a annoté 552 comme des phrases qui exprime au moins une relation de contrôle.

<i>Règle</i>	<i>Phrases annotées</i>	<i>Précision</i>	<i>Rappel</i>
R1	315	0,83492063	0,78507463
R2	158	0,65822785	0,57142857
R3	79	0,34177215	0,2967033

Tableau 3. Évaluation de trois premières règles

Ces résultats sont préliminaires, car ils correspondent à un document qui faisait partie du corpus d'où les marqueurs linguistiques ont été prélevés. Cependant, il a servi pour repérer des sources d'erreurs dans le traitement. Nous avons identifié quatre sources d'erreurs : 1) la segmentation des énumérations, très nombreuses dans les textes de SBI, 2) l'annotation des participes statiques par TreeTagger, qui affecte principalement les résultats de la troisième règle, et 3) la présence des expressions modalisées avec un adverbe (*peut ensuite modifier*). Ces trois sources correspondent au trois types d'actions correctifs qui pourraient améliorer les résultats du processus d'annotation : l'identification des bugs dans le processus, l'amélioration des résultats de TreeTagger via un post traitement, et la reformulation des règles d'EC à partir des résultats expérimentaux, en particulier celles où l'indice verbal n'est pas assez précis (règle 3) et nécessite, peut-être, la prise en compte de certaines prépositions.

6. Conclusion et perspectives

Nous avons présenté ici une méthode d'annotation automatique des relations de contrôle dans les textes de SBI. Une analyse linguistique basée sur la sémantique discursive a été faite et neuf règles d'exploration contextuelle pour l'annotation des relations de contrôle ont été produites. Ces règles se fondent sur les marqueurs les plus pertinents pour le repérage du contrôle : les verbes, et en particulier, les verbes évolutifs qui renvoient à des actions. Les règles ont été implémentées dans le moteur d'annotation EXCOM et une évaluation des trois premières règles a été présentée. Dans le travail en cours, les neuf règles devront être testées et évaluées dans un corpus de textes de SBI industrielles, et les résultats aideront à valider des modèles utilisés dans la conception de logiciels, comme le modèle d'activité ou de collaboration.

Références

- BEN-ACHOUR C. (1999). *Extraction des besoins par analyse de scénarios textuels*. Thèse de Doctorat, Université de Paris 6.
- DESCLÉS J.-P. (1990). *Langages applicatifs, langues naturelles et cognition*. Hermès, Paris.
- DESCLÉS J.-P., JOUIS C., OH H.-G., MAIRE-REPPERT D. (1991). « Exploration contextuelle et sémantique : un système expert qui trouve les valeurs sémantiques des temps de l'indicatif dans un texte ». In *Knowledge modeling and expertise transfer*. IOS Press, Amsterdam.
- DESCLÉS J.-P. (1994). « Relations casuelles et schèmes sémantico-cognitifs ». In *Langages* 113 : 113-125.

- DJIOUA B., GARCÍA FLORES J., DESCLÉS J-P. (2006). « EXCOM : an automatic annotation engine for semantic information ». In *FLAIRS 2006 19th International Florida Artificial Intelligence Research Society* (à paraître).
- FANTECHI A., GNESI S., LAMI G. (2003). « Application of Linguistic techniques for use case analysis ». In *Requirements Engineering Journal* 8 (3) : 161-170.
- GARCÍA FLORES J. (2004). « Linguistic Processing of Natural Language Requirements : The Contextual Exploration Approach ». In *Actes de Tenth Anniversary International Workshop on Requirements Engineering : Foundations for Software Quality (REFSQ'2004)*.
- GERVASHI V., NUSEIBEH B. (2002). « Lightweight validation of natural language requirements ». In *Software Practice and Experience* 32 : 113-133.
- JACKIEWICZ A. (1998). *L'expression de la causalité dans les textes. Contribution au filtrage sémantique par une méthode informatique d'exploration contextuelle*. Thèse de Doctorat, Université Paris-Sorbonne.
- LE PRIOL F. (2000). *Extraction et capitalisation automatique de connaissances à partir de documents textuels. SEEK-JAVA : identification et interprétation de relations entre concepts*. Thèse de Doctorat, Université Paris-Sorbonne.
- MOURAD G. (2001). *Analyse informatique des signes typographiques pour la segmentation de textes et l'extraction automatique des citations. Réalisation des Applications informatiques : SegATex et CitaRE*. Thèse de Doctorat, Université Paris-Sorbonne.
- SAWYER P., COSH K. (2004). « Supporting MEASUR-driven analysis using NLP tools ». In *Actes de Tenth Anniversary International Workshop on Requirements Engineering : Foundations for Software Quality (REFSQ'2004)*.
- SCHMIDT H. (1994). « Probabilistic Part-of-Speech Tagging Using Decision Trees ». In *International Conference on New Methods Language Processing*. Manchester : 44-49.
- SOH J.-S. (2002). *Transitivité et ses corrélats conceptuels*. Thèse de Doctorat, Université Paris-Sorbonne.
- ZAVE P., JACKSON M. (1997). « Four dark corners in requirements engineering ». In *ACM Computing Surveys* 6 (1) : 1-30.