

## ContextO, un outil du projet FilText orientée vers le filtrage sémantique de textes

Gustavo Crispino\*, Slim Ben Hazez\*\*, Jean-Luc Minel\*\*

\* Université de la République  
J. Herrera y Reissig 565  
11300 Montevideo - Uruguay  
crispino@fing.edu.uy

\*\* CAMS/LaLic, UMR du CNRS,  
96 Boulevard Raspail  
75 006 Paris - France  
{benhazez,minel}@msh-paris.fr

---

### Résumé

Nous présentons dans cet article un modèle conceptuel de représentation des connaissances linguistiques dédiées au filtrage sémantique ainsi qu'une plate-forme logicielle ContextO qui permet d'appliquer ces connaissances. Nous avons défini un modèle conceptuel et un langage de description et de traitement des connaissances linguistiques. Ces connaissances sont gérées par un système dédié et indépendant des applications qui les utilise. Les traitements sont spécifiés sous forme déclarative dans un langage formel que nous présentons. L'architecture du système est décrite ainsi que les composants logiciels.

---

### 1 Introduction

La notion de résumé automatique devient un des grands thèmes du Traitement Automatique des Langues. En effet, il est plus facile de lire quelques lignes ou quelques pages susceptibles d'apporter l'information cherchée que de lire des centaines de pages pour s'apercevoir qu'aucune information nouvelle ne s'y trouve. Un document textuel devra donc être maintenant géré en même temps que son résumé qui sera, par ailleurs, un des moyens d'accès au contenu du document.

Les techniques statistiques fondées sur des fréquences d'occurrences et de cooccurrences [Salton 83] ont été largement utilisées avec des résultats mitigés, au moins du point de vue des utilisateurs. Cependant, depuis quelques années on détecte un certain nombre de modèles qui font appel de plus en plus aux connaissances linguistiques. Ainsi, le modèle de Marcu [Marcu 97] s'appuie sur la Rhetorical Structure Theory [Mann et Thompson 88] et sur l'analyse des connecteurs pour construire des arbres rhétoriques qui hiérarchisent l'importance des parties textuelles. De son côté, Lehman [Lehman 95] repère des fragments textuels sur la base de scores calculés pour chaque phrase en fonction de termes préétablis. Quant à Berri [Berri 96A, Berri et alii 96B], il cherche à attribuer des étiquettes sémantiques à certaines phrases de façon à les sélectionner ou non dans un résumé. Masson [Masson 98] reconnaît partiellement des structures thématiques dans un texte, alors que Ellouze [Ellouze 98] exploite différents types d'objets textuels pour produire des schémas de résumé.

Les évaluations réalisées sur certains systèmes [Minel et alii 97 ; Jing et alii 98] ainsi que les travaux menés en collaboration avec les résumés professionnels [Endres-Niggemeyer 93] ou en comparaison avec les résumés produits par ces professionnels [Saggion et Lapalme 98] ont néanmoins montré la difficulté à réaliser des résumés standard, c'est-à-dire construits sans tenir compte des besoins des utilisateurs. La première réalisation que nous avons menée avec le projet SERAPHIN en partenariat avec la Direction des Études et des Recherches (DER) d'EDF [LeRoux 94] était fondée sur des connaissances linguistiques. Elle visait une application précise : la diffusion sélective d'informations par l'interrogation en ligne d'une base de données textuelles de façon à fournir un produit intermédiaire entre d'un côté, la chaîne d'indexation, trop pauvre, ambiguë et incapable de donner un aperçu complet du texte et d'un autre côté, le texte intégral, trop lourd à manipuler par les utilisateurs car trop riche et pas assez sélectif.

Cette expérience du résumé automatique acquise par la réalisation de la maquette SERAPHIN nous a amené à élargir le champ de nos recherches en visant non plus de simples résumés automatiques non ciblés mais des *systèmes automatiques de filtrage d'informations extraites de textes*, à l'aide de critères sémantiques qui donnent à un utilisateur la possibilité de définir son *profil de filtrage* en fonction de son objectif. Ces systèmes font appel essentiellement à des connaissances linguistiques générales, c'est-à-dire indépendantes des domaines particuliers (chimie, ingénierie, pharmacie,

médecine...). Ces connaissances s'expriment sous forme de marqueurs discursifs explicites (morphèmes, mots, expressions et locutions...) caractéristiques d'une intention pragmatique de l'auteur du texte, intention que le système doit être capable d'identifier et d'interpréter sémantiquement en fonction du contexte, comme le fait le lecteur ou résumeur peu averti du domaine traité.

Les systèmes de filtrage d'informations s'appuient d'une part, sur la méthode d'exploration contextuelle qui en constitue la justification théorique et d'autre part, sur la plate-forme logicielle ContextO qui met en œuvre les connaissances linguistiques nécessaires au filtrage sémantique.

La plate-forme logicielle ContextO est le fruit d'une collaboration entre l'Université de la République (Uruguay) et l'équipe LaLIC du CAMS (UMR 8557 du CNRS - EHESS - Université Paris-Sorbonne) et a reçu le soutien du programme ECOS-Uruguay (Action n° U97E01).

## 2 Principes et objectifs de FilText

Le concept de plate-forme d'ingénierie linguistique dédiée au traitement textuel conçue comme « une boîte à outils » [Herviou 96 ; Meunier 98] propose de définir un modèle conceptuel général de représentation de certaines connaissances linguistiques, puis de développer des tâches spécialisés qui coopèrent entre elles. Le projet FilText reprend en partie ce paradigme et s'appuie

sur la méthode d'exploration contextuelle [Desclés 91, 96, 97]. Cette méthode présente l'avantage d'une part, de rendre le travail linguistique indépendant de son implémentation informatique et d'autre part, d'articuler effectivement dans une même architecture logicielle les deux types de travaux. Dans cette approche, les linguistes analysent les textes en identifiant et en capitalisant tous les indices grammaticaux et lexicaux pertinents pour la résolution d'un problème, puis ils conçoivent et écrivent les règles d'exploration du contexte de ces indices identifiés dans un texte. Cette méthode n'est donc pas limitée à des traitements spécifiques mais elle offre un cadre de travail réaliste comme le souligne M. Charolles [Charolles 98] :

*« La délimitation des univers d'énonciation mettant en jeu, comme on le verra sur quelques exemples, une pluralité d'indicateurs, je plaiderai en faveur d'une approche du type « exploration contextuelle »... Je montrerai que cette approche offre aux linguistes non pas un modèle général d'interprétation des discours – objectif à mon avis impossible à atteindre et qu'elle ne prétend du reste pas atteindre – mais un cadre théorique réaliste aussi bien pour le développement d'applications pratiques comme le filtrage ou le résumé automatique de textes que pour celui d'hypothèses psycholinguistiques locales, hypothèses dont certaines ont été encore peu explorées. »*

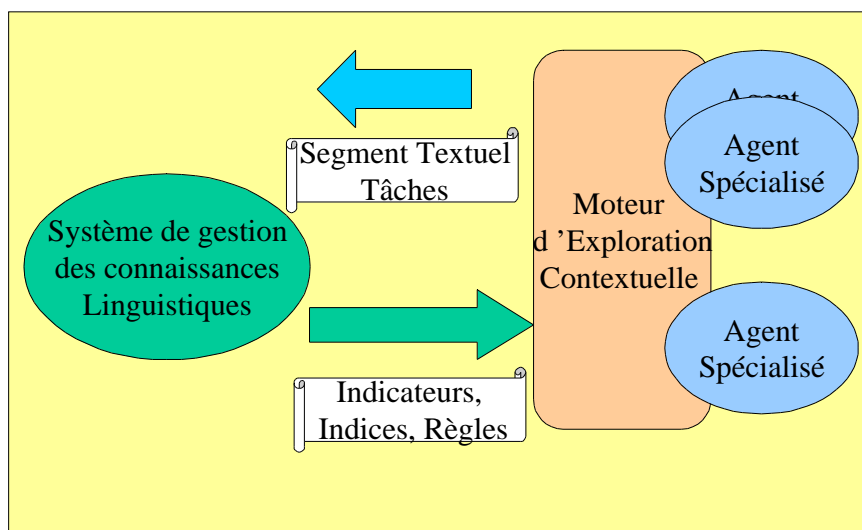


Figure 1 : Architecture conceptuelle du projet Filtext

Nous avons défini un modèle conceptuel et un langage de description de ces connaissances que nous décrivons dans la section 3.1.1. Ces connaissances sont gérées par un système indépendamment des applications qui les utilise, cela en vue d'assurer une pérennité, une capitalisation et une réutilisabilité de ces connaissances. Les règles d'exploration du contexte sont écrites dans un langage formel, indépendant de tout langage de programmation, que nous présentons dans la section 3.1.2.

### 3 ContextO

Nous avons développé une plate-forme logicielle, appelée ContextO, qui implémente le modèle conceptuel de Filtext et qui est capable d'accueillir les connaissances linguistiques et de les exploiter, indépendamment de la plate-forme matérielle ; pour ce faire nous avons choisi de développer l'ensemble en utilisant le langage JAVA.

#### 3.1 Le système de gestion des connaissances linguistiques

La méthode d'exploration contextuelle postule qu'il est possible de repérer certaines informations sémantiques à partir de marques de surface en réponse à des besoins spécifiques d'utilisateurs cherchant à sélectionner des informations importantes, comme par exemple :

- repérer les actions dans des textes techniques [Garcia 98] ;
- repérer les relations causales entre des situations [Jackiewicz 98] ;
- identifier les définitions des termes proposés explicitement ou implicitement par un auteur et les annonces thématiques mises en évidence par un auteur [Cartier 97].

Le travail préalable du linguiste consiste alors à étudier systématiquement un corpus de textes pour y rechercher des régularités lexicales et discursives dont l'emploi est représentatif de la catégorie sémantique considérée. D'après l'hypothèse de travail qui jusqu'ici s'est révélée féconde, les modes d'expressions associés à ces catégories discursives dans les corpus, sont en nombre fini. Par conséquent, cela n'exige ni le repérage de structures syntaxiques spécifiques, comme peuvent le faire [Rebeyrolle et alii 98] qui s'appuient sur les travaux de Z. Harris, ni la construction d'ontologies du domaine en vue d'en énumérer les concepts ou plus modestement les éléments thématiques. Pour appliquer la méthode d'exploration contextuelle, le linguiste doit accumuler les unités textuelles qui sont les expressions linguistiques - les marques des catégories discursives explorées - ; ces expressions

peuvent être de simples unités lexicales comme le verbe *présenter* ou des unités composées comme *les lignes suivantes*. Ces marqueurs linguistiques lorsqu'on les identifie dans un texte sont des indices (au sens de C.S. Peirce) de ces catégories. Certains indices sont plus importants que d'autres, ce sont les indicateurs. Le linguiste doit ensuite spécifier les dépendances contextuelles entre ces indices.

#### 3.1.1 Le langage de description des connaissances linguistiques

Nous avons défini un langage de description qui permet au linguiste de constituer sa base de connaissance en spécifiant : les tâches, les indicateurs ou les indices pertinents et les règles d'exploration contextuelles associées. Plutôt que de présenter ce langage sous la forme d'une grammaire formelle nous allons illustrer son utilisation par quelques exemples.

Forme	Nom de Classe
maintenant	&partie_document1
ci-dessus	&partie_document1
chapitre	&partie_document2
lignes	&partie_document2
présente	&verbe_présentatif
présentons	&verbe_présentatif
présenterais	&verbe_présentatif
présenterons	&verbe_présentatif
Il + &être + &importance	&soulignement

**Table 1** : Déclaration des formes

Dans la table 1, le linguiste déclare soit des formes lexicales significatives (des indices ou des indicateurs) qu'il organise en classes<sup>1</sup> non nécessairement disjointes, soit des combinaisons de classes (dernière ligne du tableau). Ces combinaisons permettent de déclarer des lexies ; ainsi la déclaration de *il + &être + &importance* permet au linguiste de déclarer des lexies du type *il est primordial ; il est particulièrement important ; il est, ..., essentiel ; etc.* Bien que des outils d'aide à l'acquisition permettent de produire automatiquement toutes les formes fléchies ou dérivées, il convient de remarquer que le travail du linguiste consiste justement à ne retenir que certaines formes fléchies, car pour une tâche donnée seules certaines flexions d'un verbe peuvent être significatives. Par exemple, dans le but de rechercher les annonces thématiques d'un article scientifique, le verbe *présenter* est seulement significatif lorsqu'il est employé à l'indicatif présent

<sup>1</sup> Chaque nom de classe est identifiée par un nom précédé du caractère &.

ou au futur, à la première personne du singulier ou du pluriel.

Dans la table 2, ces formes sont déclarées comme des indices complémentaires ou des indicateurs pertinents qui peuvent être associées à une ou plusieurs tâches.

Nom de Classe	Type	Nom de Tâche
&partie_document1	indice	résumé
&verbe_présentatif	indicateur	résumé
&partie_document2	indicateur	résumé

**Table 2 :** Déclaration des indices et indicateurs

Une tâche a pour finalité de regrouper des règles d'exploration contextuelle comme le montre la table 3 et correspond généralement à un processus d'étiquetage sémantique d'un segment textuel précisé.

Nom de la Règle	Etiquette attribuée	Nom de Tâche	Nom de Classe
RC1	Thématique_2	résumé	&partie_document2
RC2	Thématique_2	résumé	&verbe_présentatif

**Table 3 :** Déclaration du ou des noms de règles

Par exemple, la première ligne de la table 3 déclare une règle de nom *RCenthe1001* de la tâche *résumé*, déclenchée par une occurrence, dans une phrase du texte à analyser, d'un indicateur de la classe *&partie\_document2*. Cette règle attribue l'étiquette *Thématique\_2* à la phrase considérée. D'une manière plus générale, une règle peut déclencher différents types de décision. Des outils d'aide à la gestion de la cohérence et à l'intégration des connaissances issues d'autres travaux linguistiques permettent de répondre à l'objectif d'une acquisition incrémentale et capitalisable des connaissances.

### 3.1.2 Le langage des règles d'exploration contextuelle

Les règles d'exploration contextuelle sont exprimées dans un langage formel de type déclaratif. Ce langage est centré sur la notion d'un espace de recherche, c'est-à-dire sur un segment textuel déterminé à partir de l'indicateur dans lequel les indices complémentaires doivent être recherchés. L'intérêt de cette notion est qu'elle permet au linguiste de construire simplement un espace sans que celui-ci soit nécessairement formé de phrases contiguës dans le texte. Nous pensons qu'il est important de pouvoir exprimer des

contraintes qui prennent en compte la dimension textuelle.

Chaque règle comprend une partie *Déclaration d'un Espace de Recherche E*, une partie *Condition* et une partie *Action* qui n'est exécutée que si la partie *Condition* est vérifiée.

La partie *Déclaration d'un Espace de Recherche E* permet de construire un segment textuel, l'espace de recherche, en appliquant différentes opérations sur la structure du texte construite par le moteur d'exploration contextuelle. Il est possible de construire plusieurs espaces de recherche dans une même déclaration. Une dizaine d'opérations ont été définies pour construire un espace de recherche à partir de la structure d'un texte.

**Nom de la règle :** Rsoulignement ;  
**Tâche déclenchante :** Thématique ;  
**Commentaire :** capte un schéma du type : *il semble .. crucial*  
**Classe de l'Indicateur :** &modal3;  
**E1 :** Créer\_espace(PhraseParent\_de\_Indicateur);  
**L2 :** &adjectif-nécessité  
**Condition :** Il\_existe\_un\_indice z appartenant\_à E1 tel\_que  
 classe\_de z appartient\_a (L2) ;  
**Precéder** (Indicateur, z, contrainte) ;  
**Actions :**  
**1 :** Attribuer(PhraseParent\_de\_Indicateur  
 « Soulignement\_Auteur » )

**Figure 2 :** Un exemple de règle écrite dans le langage formel.

La partie *Condition* explicite les conditions que doivent vérifier les indicateurs et les indices complémentaires. Le langage actuel permet d'exprimer différentes conditions, comme l'existence, la position et l'agencement (*précéder, suivre, distance*) des indices. D'autres conditions permettent d'exprimer des contraintes sur les attributs des unités lexicales ou sur les morphèmes qui les composent.

La partie *Action* indique le type d'actions réalisées par la règle. Actuellement les actions possibles sont : attribuer une étiquette à un segment textuel ou déclencher une autre tâche.

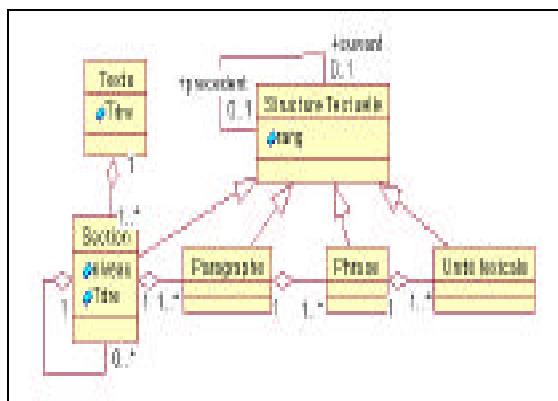
## 3.2 Le moteur d'exploration contextuelle

Le moteur d'exploration contextuelle exploite les connaissances linguistiques pour une ou plusieurs tâches choisies par l'utilisateur. Il est composé de deux systèmes qui coopèrent.

### 3.2.1 L'analyseur de texte

L'analyseur de textes a pour objet de construire une première représentation qui reflète

l'organisation structurelle du texte. Il s'appuie pour cela sur le texte balisé par un segmenteur développé par G. Mourad [Mourad 99]. Ce segmenteur applique des règles heuristiques pour reconnaître les sections (avec leurs titres), les paragraphes, les phrases et les citations. L'analyseur peut ainsi construire une structure hiérarchique illustrée par la figure 3 (formalisme UML) qui est utilisée par les opérations spécifiées dans les règles d'exploration contextuelle.



**Figure 3 :** La structure hiérarchique (simplifiée) d'un texte.

Ce modèle est ensuite enrichi par les agents spécialisés par des structures qui modélisent les chaînes de liage, segments textuels, cadres de discours [Charolles 98, Adam 90], etc., en vue d'améliorer la cohésion et la cohérence textuelle des extraits.

### 3.2.2 L'exécuteur

L'exécuteur déclenche pour toutes les tâches choisies par l'utilisateur les règles associées à celles-ci. Les règles sont considérées comme indépendantes par conséquent l'ordre de leur déclenchement, pour une tâche donnée, est indifférent. Nous pensons d'une part que ce principe est plus facilement maîtrisable lors de l'écriture des règles par un non-spécialiste, en évitant notamment les problèmes de manipulation d'arbres de décision, et d'autre part il correspond mieux à l'hypothèse que certaines marques sémantiques ne sont pas exclusives entre elles. En d'autres termes, la présence d'une négation dans une phrase conclusive n'implique pas que cette phrase n'est pas par ailleurs une information conclusive. L'exploitation de ces informations éventuellement conflictuelles est de la responsabilité des agents spécialisés. Toutes les déductions effectuées par les règles sont attribuées aux éléments qui composent la hiérarchie du texte et produisent ainsi une structure hiérarchisée « décorée » par des informations sémantiques (figure 4). Enfin, il convient de noter que le langage formel de déclaration des règles ne présume en rien

des outils informatiques (automates, moteur spécialisé, code dédié, etc.) utilisés pour les implémenter.

### 3.3 Les agents spécialisés

Les agents spécialisés ont pour tâche d'exploiter les « décorations sémantiques » du texte en fonction des besoins définis par l'utilisateur. Il existe ainsi un agent résumeur qui construit un résumé composé de phrases qui correspondent à un profil<sup>2</sup> type et un agent filtreur qui construit différents extraits de textes en fonction de profils choisis par l'utilisateur. Ces deux agents exploitent les connaissances des systèmes SERAPHIN [Le Roux 94] et SAFIR [Berri 96] et proposent des interfaces qui permettent de naviguer entre le résumé ou les extraits, et le texte source. Un agent orienté vers le repérage des relations statiques est en cours de développement [Lepriol 99]. Les agents spécialisés permettent ainsi de développer des traitements spécifiques pour un utilisateur tout en exploitant le modèle générique de traitement des connaissances linguistiques, ils se rapprochent en cela de la notion d'intelliciel développé au LANCI [Meunier 98].

## 4 Conclusion

La plate-forme logicielle ContextO est actuellement opérationnelle et les connaissances linguistiques [Cartier 97, Garcia 98, Jackiewicz 98, Jouis 93] issues de systèmes antérieurs seront intégrées progressivement. La base de connaissances linguistiques contient environ 18 500 marqueurs et 500 règles d'exploration contextuelle. Les performances de traitement sont de l'ordre d'une page à la seconde avec une limite supérieure d'environ 250 pages, sur une plate-forme matérielle de type micro-ordinateur. Nous travaillons actuellement à l'intégration d'autres agents spécialisés ainsi qu'à la possibilité d'exploiter des textes annotés pour pouvoir exploiter les informations de type morpho-syntaxiques. Enfin le développement d'une palette de navigation textuelle qui permettra à l'utilisateur de naviguer entre les représentations (extraits, graphes) fournies par les agents spécialisés et le texte source est en cours de développement<sup>3</sup>.

<sup>2</sup> La notion de profil est présentée en détail dans (Berri 96).

<sup>3</sup> Par J. Couto sous la direction de G. Crispino.

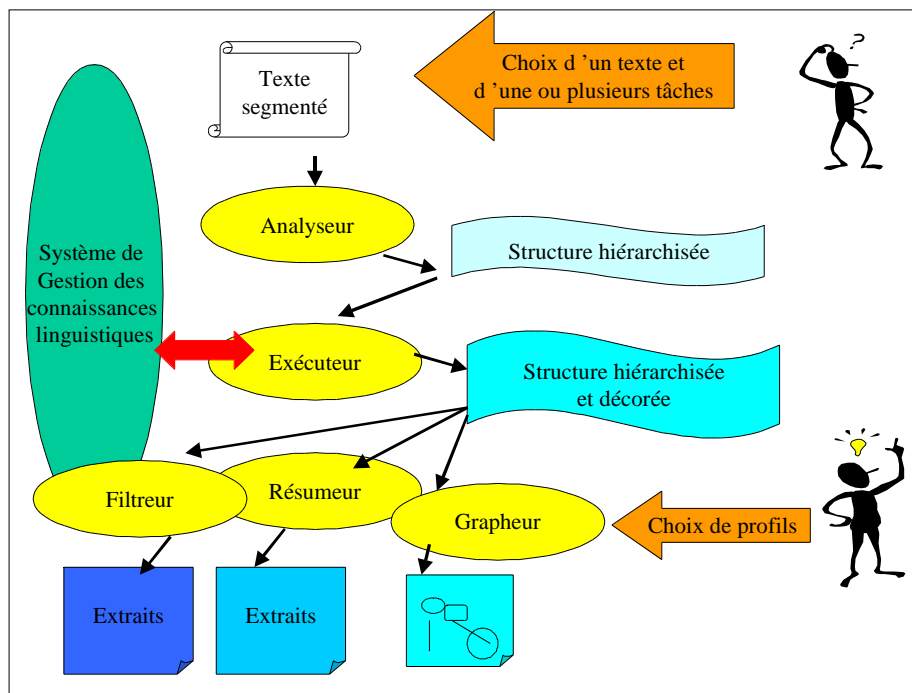


Figure 4 : Les différents composants de ContextO

## Références

- [ADAM 90] ADAM, JEAN-MICHEL. (1990 ). *Éléments de linguistique textuelle*, Mardaga, Liège.
- [BERRI 96A] BERRI, JAWAD. (1996) *Contribution à la méthode d'exploration contextuelle. Applications au résumé automatique et aux représentations temporelles. Réalisation informatique du système SERAPHIN*. Thèse de doctorat, Université Paris-Sorbonne, Paris.
- [BERRI 96B] BERRI, JAWAD, EMMANUEL CARTIER, JEAN-PIERRE DESCLES, AGATA JACKIEWICZ, JEAN-LUC MINEL. (1996). SAFIR, système automatique de filtrage de textes, *Actes du colloque TALN'96*, Marseille.
- [CARTIER 97] CARTIER, EMMANUEL. (1997). La définition dans les textes scientifiques et techniques : présentation d'un outil d'extraction automatique des relations définitoires. *TIA'97*, Toulouse,.
- [CHAROLLES 98] CHAROLLES, MICHEL. (1988). Les plans d'organisation textuelle ; période, chaînes, portées et séquences , *Pratiques*, n° 57, pp 3-13.
- [CHAROLLES 98] CHAROLLES, MICHEL. (1998). L'organisation du texte, le filtrage et le résumé. *RIFRA'98, Rencontre Internationale sur l'Extraction le Filtrage et le Résumé Automatiques*. Sfax, Tunisie.
- [DESCLES 91] DESCLES, JEAN-PIERRE, CHRISTOPHE JOUIS, HUM-GHUM OH, DANIELE MAIRE REPERT. (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*, pp.371-400, D. Herin-Aime,
- R. Dieng, J-P. Regourd, J.P. Angoujard (éds), Amsterdam.
- [DESCLES 97] DESCLES, JEAN-PIERRE, EMMANUEL CARTIER, AGATA JACKIEWICZ, JEAN-LUC MINEL. (1997). Textual Processing and Contextual Exploration Method. In *CONTEXT'97*, Rio de Janeiro, Brésil.
- [DESCLES 97] DESCLES, JEAN-PIERRE. (1997). *Systèmes d'exploration contextuelle. Co-texte et calcul du sens*. (ed Claude Guimier), Presses Universitaires de Caen, pp 215-232.
- [ELLOUZE 98] ELLOUZE, MARIEM, ABDELMAJID BEN HAMADOU. (1998). Utilisation de schémas de résumés en vue d'améliorer la qualité des extraits et des résumés automatiques. *RIFRA'98, Rencontre Internationale sur l'Extraction le Filtrage et le Résumé Automatiques*. Sfax, Tunisie.
- [ENDRES-NIGGEEMER 93] ENDRES-NIGGEEMER, BRIGITTE. (1993). An empirical process model of abstracting. In *Workshop on Summarizing Text for Intelligent Communication*, Dagstuhl, Germany.
- [GARCIA 98] GARCIA, DANIELA. (1998). Analyse automatique des textes pour l'organisation causale des actions. Réalisation du système informatique COATIS. Thèse de Doctorat, Université Paris-Sorbonne.
- [HERVIOU 96] HERVIOU, MARIE-LUCE, RICHARD QUATRAIN, MARIE-GAËLLE MONTEIL (1996). Construction de terminologies : une chaîne de traitement supportées par un atelier intégrant outils linguistiques et statistiques. *TALN'96*, Marseille.

- [JACKIEWICZ 98] JACKIEWICZ, AGATA. (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.
- [JING 98] JING, HONGYAN, REGINA BARZILAY ET KATHLEEN MCKEOWN. (1998). Summarization evaluation methods : Experiments and analysis. In *Symposium on Intelligent Text Summarization ACL*, Stanford, CA.
- [JOUIS 93] JOUIS, CHRISTOPHE. (1993). *Contribution à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes*. Thèse de doctorat, EHESS, Paris.
- [LEHMAM 95] LEHMAM, ABDELRAFI. (1995). *Le résumé de textes techniques et scientifiques, aspects linguistiques et computationnels*. Thèse de doctorat, Université de Nancy 2.
- [LEPRIOL 99] LE PRIOL FLORENCE (1999), A data processing sequence to extract terms and semantics relations between terms, *10<sup>th</sup> mini EURO Conference "Human Centered Processes" HCP'99*.
- [LEROUX 94] LEROUX, DOMINIQUE, JEAN-LUC MINEL, JAWAD BERRI. (1994). SERAPHIN project. *First European Conference of Cognitive Science in Industry*. Luxembourg.
- [MANN 88] MANN, WILLIAM C, SANDRA A. THOMPSON. (1988). Rhetorical Structure Theory : Toward a functional theory of text organization. *Text*, 8(3):243-281.
- [MARCU 97] MARCU, DANIEL. (1997). From discourse structures to text summaries. In *Workshop Intelligent Scalable Text Summarization ACL*, Madrid, Spain.
- [MASSON 98] MASSON, NICOLAS. (1998). *Méthodes pour une génération variable de résumé Automatique : Vers un système de réduction de textes*. Thèse de Doctorat, Université Paris-11.
- [MEUNIER 98] MEUNIER, JEAN-GUY. (1998). La gestion des connaissances et les intelligents. *RIFRA'98, Rencontre Internationale sur l'Extraction le Filtrage et le Résumé Automatiques*. Sfax, Tunisie.
- [MINEL 97] MINEL , JEAN-LUC, SYLVAIN NUGIER, GERALD PIAT. (1997). How to appreciate the Quality of Automatic Text Summarization. *Workshop Intelligent Scalable Text Summarization, EACL*, Madrid, Spain.
- [MOURAD 99] MOURAD, GHASSAN. (1999). La segmentation des textes par l'étude de la ponctuation. *CIDE'99*, Damas, Syrie.
- [REBEYROLLE 98] REBEYROLLE, JOSETTE ET MARIE-PAULE PERY-WOODLEY. (1998). Repérage d'objets textuels fonctionnels pour le filtrage d'information : le cas de la défintion. *RIFRA'98, Rencontre Internationale sur l'Extraction le Filtrage et le Résumé Automatiques*. Sfax, Tunisie.
- [SALTON 83] SALTION, G.M. GILL, (1983). *Introduction to Modern Information Retrieval* ; Mac Graw Hill Book Co, New York..
- [SAGGION 98] SAGGION, HORACCIO ET GUY LAPALME. (1998). Where does information come from ? Corpus Analysis for Automatic Abstracting. *RIFRA'98, Rencontre Internationale sur l'Extraction le Filtrage et le Résumé Automatiques*. Sfax, Tunisie.