

Réutilisabilité de ressources de la sémantique lexicale pour la génération de réponses en langage naturel

Florence PUGEAULT* et Guy LAPALME**

Résumé - Abstract

L'extraction de connaissances et la production de résultats en langue naturelle représentent un enjeu majeur en informatique linguistique, tant au niveau des ressources que des applications. Cet article montre les avantages de l'utilisation des ressources développées pour un prototype d'analyse dans le contexte de la génération d'énoncés. Nous présentons GASPARG, un générateur de réponses résumantes suite à des requêtes de la direction d'une entreprise sur l'état de ses projets internes. GASPARG exploite les résultats et les ressources de PAPINS, un prototype d'analyse pour l'extraction d'index structurés. Nous montrons ici que la sémantique lexicale est un outil puissant pour la génération et que les ressources linguistiques utilisées pour l'analyse sont appropriées pour la production de réponses.

Knowledge extraction and natural language utterances generation are main issues in computational linguistics either at the resources level or at the application one. This paper demonstrates, in a generation context, the advantage of using linguistic resources originally developed for an interpretation prototype. We present GASPARG, which produces synthetic responses for answering queries about the internal projects of a company. GASPARG uses the results and the resources of the PAPINS prototype dedicated to the structured index extraction. We show that lexical semantics is a powerful tool for the generation and that linguistic resources used for interpretation are appropriate for the production of responses.

Mots clefs - Keywords

Sémantique lexicale, indexation structurée, génération de résumés, rôles thématiques.

Lexical semantics, structured indexing, synthesis generation, thematic roles.

* CISI, 13 rue André Villet, B.P. 4042, 31029 CEDEX 4, Toulouse, France.

** Département d'informatique et de recherche opérationnelle Université de Montréal, C.P. 6128 Succursale Centre Ville, Montréal, Québec, Canada, H3C 3J7.
lapalme@iro.umontreal.ca

1. Introduction

Cet article* présente GASPAR (Générateur Automatique de Synthèses à partir de formes Prédicat-Arguments et Rôles thématiques), un prototype qui produit des réponses intelligentes aux requêtes d'un utilisateur sur les activités d'une grande entreprise. Notre corpus contient les représentations sémantiques de rapports techniques d'activités de recherches, les *Actions de Recherche et de Développement* (ARD) qui sont des représentations produites par PAPINS (Prototype d'Analyse pour la Production d'Index Sémantiques) développé à l'Institut de Recherche en informatique de Toulouse (IRIT) en collaboration avec la *Direction des Études et Recherches d'Électricité de France* (DER-EDF) (Pugeault F. 1995). C'est un prototype d'analyse qui transforme les ARD en structures prédicatives. GASPAR est un système de génération fondé sur les résultats de PAPINS pour fournir des réponses à des questions de veille technologique.

Notre problématique est originale en deux points : d'une part elle se fonde sur la réutilisabilité partielle des ressources inhérentes à l'extraction de connaissances et d'autre part, les réponses produites par GASPAR ne sont pas les phrases originales des ARD, mais elles sont plutôt re-générées à partir des représentations des ARD par PAPINS.

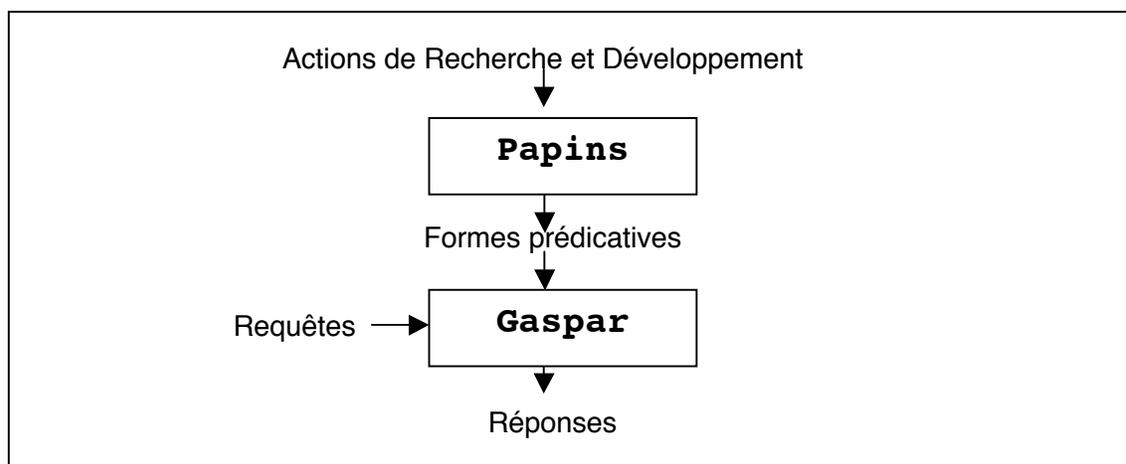


Figure 1 : Architecture globale de l'application

Comme le montre la figure 1, le système PAPINS transforme chaque phrase des ARD en structures prédicat-arguments où les arguments sont étiquetés par

* Ce projet a été réalisé au Département d'informatique et de recherche opérationnelle de l'Université de Montréal dans le cadre d'un stage post-doctoral financé par l'AUPELF-UREF.

des rôles thématiques qui sont ensuite utilisés par GASPARD pour générer des réponses à des requêtes de l'utilisateur.

Dans cet article, suite à une brève présentation de PAPINS, nous mettons l'emphase sur GASPARD en spécifiant les données nécessaires ainsi que les algorithmes mis en œuvre pour l'identification du contenu et la production des réponses en langage naturel. Nous montrons la génération de réponses par l'exploitation de rôles thématiques et l'utilisation des différents centres d'intérêts des phrases. Enfin, nous comparons brièvement notre approche avec certains travaux connexes.

2. PAPINS

PAPINS analyse des ARDs qui sont des bilans de recherche rédigés par des ingénieurs d'EDF dans le but de répondre à la question " Qui fait quoi dans l'entreprise ? " tel celui-ci (les noms ont été changés ici) :

Le groupe G1 s'intéresse plus particulièrement à l'indexation automatique. Des travaux avec CHM ont été développés dans le cadre du projet GASP. GASP regroupe des partenaires extérieurs comme G2, G3 et G4.

Nous avons étudié 6 types de questions portant sur les sujets suivants:

1. Importance des outils et des techniques utilisées
2. Diffusion ciblée
3. Synergies entre projets
4. Relations entre les partenaires extérieurs
5. Synthèses ou résumés de projets
6. Historiques de projets.

PAPINS produit des représentations sémantiques partielles des connaissances des ARD. Il y a deux niveaux d'analyse qui correspondent respectivement à deux types de représentations sémantiques de différents degrés de granularité (Pugeault F *et al.* 1994) :

- Le niveau **pragmatique** représente l'organisation externe des textes et donc la manière dont les phrases des textes sont rhétoriquement reliées entre elles. Il produit des fragments de phrases pertinents organisés en *articulations* : thème, motivations, problèmes et réalisations. Ce niveau indique ce qui est traité globalement dans les ARD.
- Le niveau **linguistique** extrait des formes prédicatives dans lesquelles un rôle thématique est automatiquement assigné à chaque argument. Par exemple, dans les structures prédicatives du tableau 1, on assigne les

rôles ae (agent), tg (thème), tib (bénéficiaire), pos-abs (position abstraite) et acp (accompagnement). Ce niveau produit une représentation sémantique partielle qui simule une compréhension partielle des textes.

Nous décrivons maintenant les deux principales notions linguistiques requises dans PAPINS et utilisées dans GASPAR.

2.1 Structures prédicatives

Les structures prédicatives de PAPINS contiennent différentes sortes d'informations : le nom de l'articulation associée au fragment de phrase et la représentation sémantique de l'articulation. Cette représentation contient le nom du prédicat principal, et les rôles thématiques associés aux arguments proches et aux modificateurs qui peuvent eux-mêmes être prédicatifs ; le tableau 1 en montre quelques exemples. Dans PAPINS, on définit une hiérarchie de rôles thématiques qui forment une base fiable pour une sémantique descriptive des relations entre arguments et prédicats. Nous montrons maintenant via GASPAR pourquoi les rôles thématiques sont fondamentaux dans la sélection du contenu des réponses et comment ils sont utilisés pour produire des phrases bien formées en français.

Langue naturelle	Structure prédicative
<i>Quels sont les partenaires du groupe G1 ?</i>	ques(4, [tib,acp] [quels,sont,les,partenaires,du, groupe, g1], [partenariat,[tib,[avec,[acp,g1]]]]).
<i>G1 s'intéresse à l'indexation</i>	interesser(tib, [g1], pos-abs, [par,[pos,indexation]]).
<i>G1 travaille sur GASP en collaboration avec CHM.</i>	developper(tg, [gasp], modV,[a,[loc_spat,[g1]]], modV,[en,[acp,[la_collaboration, [acp,[avec,[acp,[chm]]]]]]]).
<i>G6 est partenaire de G1 sur PAP</i>	partenariat(ae, [g6], tib,[avec,[acp,[g1]]], pos_abs,[sur,[pap]])

Tableau 1 : Exemple d'une structure prédicative reçue en entrée par GASPAR avec sa paraphrase en langue naturelle. On montre ensuite les

représentations de phrases extraites par GASPARE et la phrase en langue naturelle générée automatiquement.

2.2 Le thesaurus EDF

Le thesaurus EDF contient environ 20 000 termes qui sont hiérarchiquement organisés en champs sémantiques eux-mêmes structurés en thèmes sémantiques. Les relations sémantiques les plus couramment exprimées dans ce thesaurus sont la relation “ générique/spécifique ”, la relation “ voir aussi ” et la relation “ est utilisé pour ”. Dans PAPINS, le thesaurus a été enrichi par l’ajout de restrictions de sélection d’un niveau de granularité plus générique, de manière à ce que l’on puisse établir une chaîne menant des informations les plus spécifiques aux plus génériques. PAPINS utilise ainsi les restrictions de sélection comme un critère d’assignation de rôles thématiques. Nous allons voir en section 3.1 que GASPARE, en aval de PAPINS, exploite les relations du thesaurus à la fois pour la sélection du contenu des réponses et pour la sélection lexicale lors de la production langagière.

3. GASPARE

Le but de GASPARE est de produire des réponses “ intelligentes ”, ce qui s’oppose aux systèmes de recherche d’information généraux qui répondent simplement en fournissant la liste des documents pertinents dans lesquels l’utilisateur doit regarder et combiner l’information par lui-même. GASPARE répond en fonction du but communicatif lié au type de la requête traitée. En combinant l’ensemble des informations de notre corpus, GASPARE produit une courte synthèse résumant les réponses qui satisfont la requête de l’utilisateur. Par exemple, la figure 1 montre 3 réponses qui satisfont la question de l’utilisateur que GASPARE va organiser en une ou plusieurs phrases re-générées à partir des structures sémantiques. Le processus de génération de GASPARE se déroule en trois étapes:

1. appariement d’une requête avec un type prédéfini de question,
2. sélection du contenu informatif (cf. 3. 1),
3. production langagière (cf. 3.2).

L’approche de la première étape est similaire à celle qui avait été adoptée lors de la mise en œuvre des stratégies de PAPINS : la représentation sémantique d’une requête est associée à son type au moyen de déclencheurs sémantiques. Cette étape ne fait donc pas l’objet de cet article. GASPARE ne produit pas de résumés de textes : nous choisissons uniquement les informations pertinentes à retenir parmi les diverses représentations sémantiques des ARD. C’est pourquoi

la deuxième étape consiste à extraire le thème ou le centre d'intérêt de la question (que nous désignerons par *focus* dans la suite de ce texte) et à identifier les phrases du corpus qui combinent le focus principal et les critères spécifiés dans la représentation de la question (rôle thématique et articulations). Enfin, lors de la dernière étape, les représentations adéquates collectées sont utilisées pour construire des phrases bien formées par le biais des relations de correspondance entre les rôles thématiques et les fonctions grammaticales. Le focus est souvent relié à un syntagme nominal et est représenté par l'un des arguments essentiels du prédicat. Toutefois, pour améliorer la précision et a fortiori la pertinence de la sélection du contenu informatif, le focus peut être étendu à un syntagme prépositionnel voire à une structure prédicative complète. Différents focus apparaissent dans notre étude :

global: ce sur quoi la question est centrée ;

courant : le centre d'intérêt de la phrase courante ;

potentiel : un des centres d'intérêts secondaires de la phrase traitée.

Dans GASPARD, le traitement du focus va permettre d'extraire le centre d'intérêt d'une requête et de l'apparier aux centres d'intérêts des phrases des ARD. Ce processus est dynamique dans la mesure où le focus global peut être étendu au focus courant ou à un focus potentiel s'il existe. Brièvement, notre algorithme consiste à retenir d'abord les représentations des phrases qui contiennent à la fois le focus global et les critères de sélection qui sont exprimés via le type de la question. Ensuite, nous identifions le focus courant de la première phrase sélectionnée : le focus courant est un argument contenu dans la phrase traitée qui est distinct du focus global. Une fois le focus courant connu, nous extrayons les phrases le contenant sur la base du sous-ensemble des textes dont les phrases qui contenaient le focus global ont été extraites. Cette méthode assure l'homogénéité et la pertinence de nos résultats. De ce fait, les phrases extraites ne sont pas seulement centrées sur le focus global mais fournissent également des informations sur des thématiques qui lui sont connexes (foci courants).

La seule utilisation des foci dans GASPARD pour gérer la sélection des réponses est liée au fait que nos textes proviennent jusqu'à maintenant d'une unique source de documents. Dans le cas contraire, il est clair que les énoncés produits via de telles règles pourraient manquer d'une réelle cohérence : le lecteur pourrait faire de fausses implications à partir des textes, en inférant notamment des connexions où il n'en existe pas.

3.1 Spécification de la sélection du contenu informatif

Dans GASPAR, la phase de sélection du contenu informatif détermine le contenu de la structure de l'énoncé à produire : le type d'informations, leur ordre, leur segmentation en paragraphes et en phrases. Conformément aux approches classiques de Danlos (1985) et McKeown (1985), nous utilisons les notions de structure du discours et de schémas rhétoriques pour organiser nos réponses. Comme les ARD suivent une organisation régulière et rhétorique, nous spécifions des schémas pour guider la construction des énoncés qui

- 1) fournir des informations reliées au focus global identifié par l'argument nominal de la requête,
- 2) ne pas fournir plus de deux phrases de détails :
 - a) des spécifications sur le focus courant (s'il existe un focus secondaire dans la première phrase extraite via la clause 1) : les spécifications doivent être trouvées dans les textes traités avec le focus global,
 - b) accéder à un focus potentiel via les relations thesaurus,
 - c) revenir au focus global.
- 3) conclusion :
 - a) revenir au focus global pour clore le sujet traité,
 - b) extrapoler en prenant en compte un focus potentiel (relation "voir aussi")

prennent en compte la structure rhétorique des textes (ordre des phrases) et la gestion du focus. Toutefois, notre approche est plus dynamique que celle de K. McKeown parce que différents schémas peuvent être associés au même type de questions. L'élaboration des schémas est basée sur l'étude empirique de réponses potentielles à différentes questions du même type. Nous avons identifié des patrons en fonction de la fréquence des principales configurations et nous avons également considéré le but communicatif de chaque réponse. Notre schéma générique pour produire des réponses est présenté à la figure 2.

Figure 2 : Guide rhétorique général pour la construction des réponses

De plus, nous avons sélectionné trois critères pour identifier les connaissances pertinentes à la génération d'une réponse :

articulations : les thèmes principaux des ARD (focus global), les motivations principales décrites dans les ARD (objectifs), les problèmes relatés et les réalisations futures (solutions aux problèmes, perspectives, etc.) ;

rôles thématiques

arguments essentiels versus modificateurs.

3.1.1 Les articulations

Les articulations se comportent comme des prédicats rhétoriques dont les arguments sont les représentations prédicatives de PAPINS. Les articulations agissent sur le focus, favorisent la sélection d'informations générales et l'extraction des phrases qui correspondent aux synthèses globales sur certains sujets (motivations, problèmes). Ainsi, le découpage d'un texte en articulations est suffisant pour répondre à des requêtes générales.

3.1.2 Les rôles thématiques

A cause de la précision des rôles thématiques (Pugeault F *et al.* 1994), nous pouvons mettre en évidence les informations correspondantes à des sujets spécifiques : collaboration, partenariats, importance de techniques, etc. Grâce à leur organisation hiérarchique, les rôles thématiques fournissent des informations sémantiques pour produire des réponses selon différents niveaux de granularité. Par exemple, le rôle *source-abstraite* est plus spécifique que *localisation* et favorise la sélection d'informations plus pertinentes et plus précises.

3.1.3 Les types d'arguments

En distinguant les arguments essentiels des modificateurs, nous établissons différents niveaux de précision des réponses. Les modificateurs sont très spécialisés et dénotent de ce fait une information spécifique, alors que les arguments essentiels relatent un niveau d'information plus général. D'un autre côté, les modificateurs contiennent des informations annexes qui peuvent être non pertinentes, alors que les arguments essentiels sont reliés au focus global de la requête : les questions telles que *qui*, *quoi* ou *quand*, par exemple, sont explicitement liées aux rôles *agent* et *thème*. Par ailleurs, les informations du thesaurus mettent en évidence les relations sémantiques entre les termes suivants : généralité, spécificité, synonymie et association. Ces relations enrichissent la recherche d'information: le focus est non seulement traité sous sa forme canonique, mais aussi sous ses formes reliées. Par exemple, en considérant " projet " comme focus principal, la recherche d'information utilisera aussi bien " tâche " (par synonymie) et " plan de travail " (par association). Ces relations sont utiles en génération pour la sélection lexicale parce qu'elles favorisent le choix de termes distincts pour exprimer un même concept.

3.2 La production de langage naturel dans GASPAR

Nous n'utilisons que les rôles thématiques pour produire des phrases bien formées. Pugeault (1995) explicite le fait que les rôles thématiques sont en forte

connexion avec les fonctions grammaticales. Dans GASPAR, nous spécifions des règles de correspondance entre la sémantique des rôles thématiques et la position syntaxique des arguments.

3.2.1 Rôles thématiques *versus* fonctions logiques

Beaucoup de linguistes prennent comme acquis l'association entre les rôles thématiques et les fonctions grammaticales. Cependant, des études ont montré que les rôles peuvent dépendre de la position syntaxique de l'argument auxquels ils sont alignés (Speas M. 1990). Il est commun de dire que dans " la porte s'ouvre " et " la clé ouvre la porte ", les sujets sont respectivement un thème et un moyen. Dans un tel cas, Fillmore (Fillmore C. 1968) stipule que même si un thème et un moyen peuvent être sujet, le thème ne peut pas être sujet si thème et moyen sont co-occurents. De ce fait, il suggère, comme bon nombre d'autres linguistes, qu'il existe une hiérarchie de rôles thématiques en fonction de leur comportement syntaxique. Dans ce cas, les moyens sont plus élevés dans la hiérarchie que les thèmes. Toutefois, il n'existe aucun consensus actuellement ni sur la définition des rôles thématiques, ni sur leur organisation hiérarchique. Le tableau 2 donne les principales hiérarchies thématiques qui s'approchent de nos travaux ; A < B signifie que A est pressenti comme sujet par rapport à B et A | B signifie que A et B sont au même niveau de préférence .

(Fillmore C. 1968)	agent < expérimentateur < instrument < objet < source < but < localisation < temps
(Jackendoff R. 1972)	agent < localisation source but < thème
(Dowty D. 1991)	agent < instrument < expérimentateur < patient < source but

Tableau 2 : Différentes hiérarchies thématiques proposées

Dans PAPINS, nous avons adopté le schéma suivant

agent < (thème | identifieur) < accompagnement < moyen < localisation

3.2.2 Typologie et mise en oeuvre des règles de réécriture

Pour transformer les représentations de PAPINS en formes syntaxiques de base, seules les informations sur les prédicats sont requises : les rôles thématiques, en fonction de leur ordre dans les grilles thématiques, et les prépositions, perçues comme des marqueurs d'objets indirects. Les différentes configurations grammaticales pour les différentes configurations thématiques trouvées dans notre corpus ont été énumérées pour établir des heuristiques qui

prennent en compte la distribution thématique aux arguments des prédicats. Dans PAPINS, l'ordre thématique a été utilisé pour assigner un rôle à un argument. Il était ainsi évident de l'utiliser à nouveau pour implémenter la tâche de génération. La production de langage naturel dans GASPARG se réalise en reconstituant des formes de base sans conjugaison ni accord, puis en les reformulant en phrases bien formées. Notre stratégie pour produire des formes de base est classique et consiste à : (1) rechercher un agent pour la position sujet ou chercher un thème, et s'il n'y en a pas, laisser vide la position, (2) considérer chaque syntagme nominal (SN) comme un objet direct et (3) considérer chaque syntagme prépositionnel (SP) comme un objet indirect. Pour traiter les accords et les conjugaisons, nous utilisons un lexique et une grammaire. L'étiquetage lexical développé dans PAPINS résout ce problème : il produit une liste de mots étiquetés par leur structure morpho-syntaxique. Cette liste est traitée par notre analyseur morpho-syntaxique pour reconnaître les SN et les SP.

3.2.3 Perspectives et améliorations

Nous utilisons la voix active pour la production des phrases, mais ce n'est pas toujours la meilleure solution pour une réponse. La forme passive peut aussi être utilisée pour mettre en valeur un argument objet, lorsque, par exemple, l'agent est moins important que les faits; ceci est fréquent dans les ARD parce que les agents en sont toujours les rédacteurs. Pour l'instant, GASPARG produit des phrases pertinentes sans les relier entre elles. Il serait intéressant de mixer cette information et de construire un nouveau texte discursivement cohérent et ciblé pour l'utilisateur, notamment dans le cas de gestion de documents multi-sources. On pourrait en l'occurrence prendre en compte des informations de nature temporelle et aspectuelle, ce qui permettrait de gérer acceptablement les aspects pragmatiques des textes.

Du point de vue génération, il reste à choisir quels mots sont appropriés pour produire les réponses : la sélection lexicale pour éviter la redondance des termes, le traitement des anaphores et la prise en compte de connecteurs temporels.

3.3 Vers la réutilisabilité des ressources

Nous avons décrit les ressources linguistiques requises pour la génération et nous avons vu leur appariement avec celles originalement définies pour l'extraction de connaissances dans PAPINS : le lexique, les rôles thématiques et les relations du thesaurus ont été réutilisés dans GASPARG. PAPINS est fondé sur un lexique constitué d'entrées lexicales sémantiquement riches (Pugeault F. 1994) qui sont très utiles pour produire du langage naturel. L'utilisation de

classes sémantiques de verbes combinées avec des classes sémantiques de prépositions est manifestement suffisante pour faire une bonne interprétation des ARD. Cependant, des classes générales soulèvent un important problème dans un contexte de génération : celui de la sélection lexicale. Par exemple, en considérant la classe des verbes de mouvement reliée à la classe des prépositions de lieu, comment choisir une préposition de lieu pour la combiner avec un verbe de mouvement donné ? Les récents travaux sur les classes sémantiques de verbes (Levin B. 1993 et Saint-Dizier P. 1996) traitent ce problème en se fondant sur le comportement syntaxique des verbes. Chaque configuration syntaxique de verbe est analysée et liée à sa distribution thématique et à sa liste de prépositions. Prendre en compte ces classes dans le lexique améliore son pouvoir pour la génération. Le tableau 3 récapitule les différentes ressources et leur utilisation à la fois dans PAPINS et GASPAR.

Ressources	PAPINS	GASPAR
Lexique	Étiquetage pour l'analyse	Production langagière
Grammaire	Analyse partielle	Conjugaisons et flexions
Thesaurus	Restrictions de sélection	Gestion de focus, Sélection lexicale
Rôles thématiques	Typage sémantique de concepts	Recherche d'informations Production langagière
Classes sémantiques de verbes	Critère d'assignation des rôles	Recherche d'informations, Sélection lexicale

Tableau 3 : Réutilisabilité des ressources d'analyse pour la génération.

4. Travaux connexes

Notre système est comparable à celui de (Hoffman B. 1996) qui sélectionne des résumés à partir de représentations sémantiques de textes scientifiques. Ses représentations sémantiques sont des formes prédicatives mais ne contiennent pas d'information sémantique telles que les rôles thématiques. De plus, les résumés générés sont construits à partir des phrases originales des textes, qui sont toutefois élagués d'informations imbriquées. Le système de Beryl Hoffman ne traite pas à proprement parler de génération de langage naturel, mais plutôt de sélection d'information : son système choisit ce qu'il faut préserver dans un résumé. Par ailleurs, Hoffman conserve les ambiguïtés sémantiques parce qu'elle veut respecter le style du rédacteur. D'un autre côté, GASPAR résout les ambiguïtés pour répondre par une synthèse pertinente des réponses possibles à la requête de l'utilisateur.

Le système de (Masson N. 1996) extrait les principaux sujets thématiques de textes en utilisant un marquage de surface et sans connaissance particulière du domaine traité. Cette méthode est similaire à celle adoptée dans PAPINS, mais elle est ici basée sur des critères statistiques. Comme le lecteur et le rédacteur partagent le même modèle de textes, l'effort cognitif du lecteur est ainsi réduit. Ce point est important pour la génération parce que les textes de Masson traitent de vulgarisation scientifique, cependant ce n'est pas efficace pour répondre de façon spécifique à des questions. Ce système est proche du nôtre dans la mesure où Masson analyse et formalise des phénomènes en surface des textes pour construire des représentations à partir desquelles il peut produire un résumé.

Le système SERAPHIN (Berri J. 1996), offre une méthode d'exploration contextuelle pour le résumé automatique. Cette étude focalise sur les expressions linguistiques significatives pour résoudre un problème spécifique; c'est analogue au niveau pragmatique de PAPINS. SERAPHIN extrait les phrases les plus pertinentes d'un texte pour construire un résumé qui donne une idée du contenu du texte entier. Ce n'est pas l'objectif principal de GASPAREL, mais ce point pourrait être pris en compte pour produire des réponses générales notamment lors de la création d'historiques de projets.

5. Conclusion

Cet article a montré que les rôles thématiques forment des critères pertinents pour la sélection du contenu informatif et sont des outils stables et fiables pour la production de langage naturel. Nous pouvons établir des relations entre les rôles thématiques et les fonctions grammaticales si nous traitons un domaine de langage restreint et si nous nous bornons seulement à produire des formes syntaxiques de base.

En fin de compte, nous avons montré que les ressources linguistiques requises pour l'analyse dans PAPINS sont concrètement réutilisables pour la recherche d'information et la génération de réponses en langue naturelle. Cette approche peut être implantée de façon efficace si le domaine d'application est restreint. C'est une " méthode " que nous avons proposée et la généralisation de cette approche reste bien entendu contrainte par l'effort de définition des ressources linguistiques des domaines traités. Même si les projets fondés sur les ressources de la sémantique lexicale ne sont pas encore suffisamment convaincants pour être appliqués à une grande échelle, on peut espérer à terme parvenir à adapter la réutilisation de telles ressources à des problématiques différentes comme la diffusion ciblée voire la traduction automatique.

Bibliographie

- BERRI Jawad (1996) : “ Mise en œuvre de la méthode d’exploration contextuelle pour le résumé automatique de textes. Implémentation du système SERAPHIN ”, 1^{ère} Conférence étudiante en Linguistique Informatique (CLIM), Montréal, Québec, Canada, p128-135.
- DANLOS Laurence (1985) : *Génération automatique de textes en langues naturelles*, Etudes et Recherches en informatique, Masson, Paris.
- DOWTY David (1991) : “ Thematic Proto-Rules and Argument Selection ”, *Language*, Vol 67, no 3.
- FILLMORE Charles (1968) : “ The Case for Case ”, E Bach et R. Harms ed., *Universal in Linguistic Theory*, p 1-90, Holt, Rinehart and Winston, New-York.
- HOFFMANN Beryl (1996) : “ Summarization : an application for Natural Language Generation ”, Actes 8th International Natural Language Generation Workshop, Herstmonceux Castle, Sussex, Angleterre, Poster, p. 37-40.
- JACKENDOFF Ray (1972) : “ Semantic Interpretation in Generative Grammar ”, MIT Press, Cambridge, MA., 1972.
- LEVIN Beth (1993) : *English Verb Classes and Alternations*, University Chicago Press, Chicago.
- MASSON Nicolas (1996) : “ Un système de résumé automatique de textes ”, Journées *Filtrage et résumé automatique*, CAMS-LALIC, Paris.
- McKEOWN Kathleen (1985) : *Text Generation : using Discourse Strategies and Focus Constraints to Generate Natural Language Text*, Studies in Natural Language Processing, Cambridge University Press, Cambridge.
- PUGEAULT Florence (1995) : *Extraction dans les textes de connaissances structurées : une méthode fondée sur la sémantique lexicale*, Thèse de Doctorat, spécialité Informatique, Université Paul Sabatier, Toulouse.
- PUGEAULT Florence ; SAINT-DIZIER Patrick ; MONTEIL Marie-Gaëlle (1994) : “ Knowledge Extraction from Texts : a Method for Extracting Predicate-Arguments Structures from Texts ”, Actes COmputational LINGuistics, volume 2, 1039-1043, Kyoto, Japon.
- SAINT-DIZIER Patrick (1996) : “ Constructing Verb Semantic Classes for French : Methods and Evaluation ”, Actes COmputational LINGuistics, Poster, Copenhagen, Danemark, p 1127-1130.

SPEAS Margareth (1990) : *Phrase Structure in Natural Language*, Studies in Natural Language and Linguistic Theory, Kluwer Academic Publishers.