

Apports de la Génération Automatique de Textes en Langue Naturelle à la Recherche d' Information

Laurence Balicco, Salaheddine Ben-Ali, Claude Ponton,

et Stéphanie Pouchot

ICM - Université Stendhal

Résumé

L' expérience acquise dans notre équipe en recherche d' information et en génération automatique de textes (production de textes par un système informatique à partir d' un contenu formel) nous a montré que la relation de l' utilisateur avec le système de recherche d' information n' était pas naturelle, l' utilisateur devant généralement s' adapter au système. Pour simplifier cette interaction personne-système, nous avons étudié les apports d' un système de génération automatique à un système de recherche d' information. Le processus de recherche d' information se compose de trois phases :

- formulation de la requête par l' utilisateur,
- traitement de la requête par le système,
- affichage des résultats.

En quoi et comment un système de génération automatique rend-il l' utilisation d' un système de recherche d' information plus efficace et facilite-t-il son appréhension ? Ceci fait l' objet de cette publication. Quand la formulation de la requête est inadéquate, on peut en générer automatiquement différentes reformulations en fonction des caractéristiques du système de recherche d' information. Il est également possible d' engendrer une explication sur la méthode employée et le type de données traitées par le système afin que l' utilisateur puisse comprendre le processus de recherche et adapter sa

requête. Si une trop grande quantité d' informations est délivrée, il est nécessaire de structurer, d' organiser la présentation des résultats voire, selon l' application, de produire un résumé des réponses. Le générateur intervient donc au niveau de l' interface personne-machine et communique avec le système et l' utilisateur. Nos travaux sur la structuration et la mise en forme des textes produits, sur la génération de textes explicatifs et la reformulation de requêtes ont servi de point de départ à ce travail dont l' objectif final est la mise en œuvre d' une application réelle.

Introduction

Que ce soit dans le monde de l' entreprise, dans le domaine politique, dans le cadre des loisirs ou encore de l' éducation, la communication est omniprésente (Miège 1989). L' essor des Nouvelles Technologies de l' Information et de la Communication est aujourd' hui indéniable. L' obligation actuelle de communiquer se traduit notamment par le développement exponentiel d' Internet et du nombre de ses usagers. Si en 1993, par exemple, la toile comportait quelques milliers de pages accessibles, on en compte aujourd' hui plus de 350 millions avec un taux de croissance estimé à 20 millions de pages par mois (Bourdoncle 1999). Par ailleurs, en France, l'usage d'Internet a progressé de 45% entre janvier et juin 1999¹. Dans le même temps, le nombre d' Allemands utilisant Internet chez eux a doublé durant les 5 derniers mois². De plus, le profil même des utilisateurs a évolué : les “ experts ” (ingénieurs, chercheurs, étudiants) ne sont plus les seuls à “ surfer ” . Face à ce flot d' informations, des outils de recherche (moteurs, annuaires...) permettent aux utilisateurs de chercher et, idéalement, de trouver l' information souhaitée. Comme le soulignent Jean-Paul et Marie-Christine Haton (Haton 1993), une langue naturelle est un moyen privilégié d' interaction entre l' homme et la machine. Pouvoir poser une question en langage naturel à une base de données constitue un progrès notable pour l' utilisateur, notamment au niveau ergonomique. Par ailleurs, dans ses réflexions sur l'optimisation de la recherche d'information (RI) dans les bases textuelles, Catherine Leloup souligne les

apports des techniques linguistiques par rapport aux autres techniques existantes (Leloup 1998).

En ce sens, l'expérience acquise au sein de notre équipe en RI et en génération automatique de textes nous a montré que la relation entre utilisateur et système de RI n'était pas naturelle. En effet, l'utilisateur doit le plus souvent s'adapter au système.

Dans cet article, nous présentons la génération automatique de textes en général et telle qu'envisagée dans notre équipe. Puis nous expliquons comment et à quels niveaux notre générateur peut contribuer à améliorer un processus de RI : en quoi cet outil peut-il optimiser l'utilisation d'un système de RI et faciliter son appréhension ?

1. La génération automatique de textes

1.1 Généralités

La génération de textes consiste à faire produire des textes par un ordinateur, de façon à exprimer automatiquement un contenu formel en langue naturelle. On partage traditionnellement la génération en deux étapes : le " quoi dire ? " qui consiste à déterminer le contenu du texte à engendrer, et le " comment le dire ? " qui est son expression en langue naturelle. De plus, l'objectif est l'élaboration de textes linguistiquement corrects et compréhensibles par le destinataire. En effet, l'écriture d'un texte, quelle que soit la méthode utilisée, ne se conçoit pas sans sa lecture et sa compréhension (Sfez 1999).

La qualité des textes engendrés, mais surtout leur compréhension sont fonction des informations dont le générateur dispose pour produire ces textes. Il devient alors essentiel de recenser toutes les connaissances nécessaires au générateur, afin de les lui fournir ou de lui donner les moyens de les calculer.

En effet, les textes élaborés dépendent d'abord des informations que l'on souhaite transmettre au destinataire du message. Ces éléments varient en fonction du contexte et donc de l'application visée. Ils sont représentés formellement (concepts, graphes ou autres) et constituent le " quoi dire ? " . D'autres informations plus contextuelles sont indispensables pour passer du "

quoi dire ? ” au “ comment le dire ? ” . Elles sont en partie données et en partie calculées au cours du traitement pour guider le générateur tout au long de sa tâche.

Les informations manipulées par le processus de génération sont donc de types très divers :

- générales : on parle aussi de connaissances universelles,
- sur l'application : explicites ou implicites (on choisit de détailler plus ou moins une information),
- linguistiques : essentiellement morphologiques et syntaxiques, mais aussi sémantiques,
- sur le destinataire du message, avec le principe d'un “ modèle ” décrivant le destinataire.

1.2 Le générateur de cristal

Nous avons travaillé dans l' équipe sur l' élaboration d' un générateur du français écrit, implémenté en Prolog. Cet outil est non contextuel (Balicco 1993, Ponton, 1996) c' est-à-dire indépendant de toute application. Il fonctionne sur la partie “ comment le dire ? ” et est donc fondamentalement un générateur de surface traduisant sous forme linguistique un contenu sémantique donné. Notre système est basé sur un modèle linguistique qui permet la production de plusieurs versions “ équivalentes ” d' un même contenu sémantique. Ce modèle organise le texte à l' aide d' opérations linguistiques qui servent à l' élaboration de toutes les versions différentes du texte. Nous appelons opération linguistique la mise en œuvre d' un phénomène linguistique. Deux types de phénomènes linguistiques sont considérés par notre générateur : les phénomènes de regroupement (coordinations, relatives...) (Dale 1989), (Horacek 1992) et les phénomènes référentiels (anaphores, ellipses...) (Danlos 1989), (Nogier 1991). Nous utilisons des opérations de regroupement pour produire des modifications de structure en regroupant des éléments qui peuvent être associés, entre des propositions différentes, par exemple. Cette approche est

également utilisée dans (Horacek 1992). Une opération linguistique référentielle remplace des éléments textuels par des anaphores telles que les pronoms ou les ellipses. Par exemple, on peut remplacer des syntagmes nominaux par des pronoms personnels, ce qui est un problème étudié dans différents systèmes de génération comme ceux décrits dans (Danlos 1989) ou (Nogier 1991). Qu'elles soient de regroupement ou référentielles, ces opérations doivent d'abord être détectées à partir du contenu fourni en entrée (sous forme prédictive). Le générateur passe donc en revue les différentes propositions et recherche si des opérations de regroupement (coordination par exemple) ou référentielles (telles que l'anaphorisation) sont possibles entre deux propositions. Pour cela, il a recours à des règles de détection formant le modèle linguistique et constitue ainsi un ensemble d'opérations linguistiques applicables sur le texte. Le système choisit ensuite de réaliser certaines d'entre elles. Ce choix est paramétrable et permet donc de modifier la composition linguistique des textes.

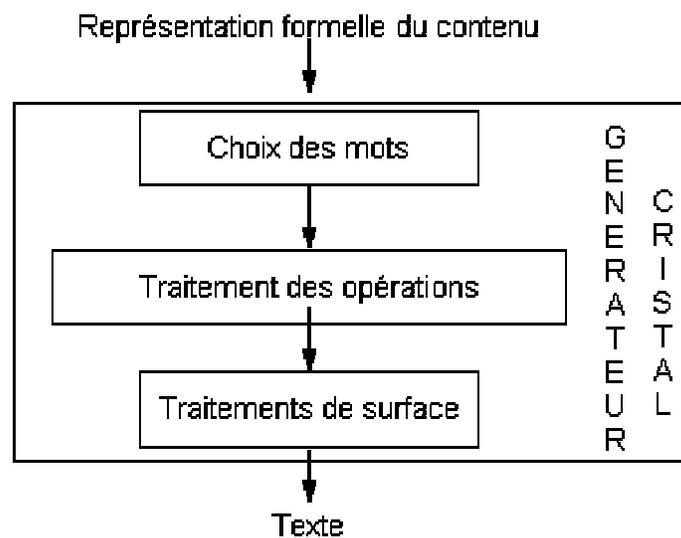


Fig. 1. Fonctionnement du générateur CRISTAL

Comme le montre le schéma ci-dessus, le processus de génération est décomposé en plusieurs tâches qui sont définies dans des modules distincts. Ils peuvent donc être modifiés indépendamment sans conséquence sur le fonctionnement général du générateur. Le système commence par le choix des mots (un message sous forme logique est transformé en unités lexicales), puis la

détermination de la structure syntaxique à l' aide des opérations linguistiques. La génération se poursuit par une phase de génération morphologique (calcul des formes fléchies, conjuguées...) avant de terminer par des traitements de surface (majuscules, élisions...). Les opérations linguistiques sont concrètement réalisées lors de l' écriture du texte. Pour cela, des règles de transformation sont appliquées à la représentation formelle du contenu du texte. Le générateur utilise des ressources linguistiques indépendantes de l' application (un lexique et une grammaire) et un lexique propre à chaque application.

Il nous a semblé important de tester et valider cet outil. Pour cela, nous lui avons associé une première application consistant à produire des textes équivalents à ceux d' un corpus en langue naturelle (Balicco 1997). Ce corpus a été recueilli par une équipe de l' INRIA³ (Bisseret, Montarnal 1996). Plusieurs raisons ont guidé ce choix.

Tout d'abord, ce corpus a été rassemblé dans le but d'étudier les processus de linéarisation⁴ et les stratégies lors de la production de textes descriptifs. Cette tâche repose sur des processus cognitifs complexes. Une étude des textes conçus dans ce cadre fournit donc des bases solides pour notre objectif de simulation automatique de ces processus (Balicco, Pouchot 1999), tout travail d'écriture nécessitant une phase préalable de linéarisation. Ce qui est observé chez les sujets humains permet de mieux appréhender la génération automatique.

Ensuite, le monde auquel il est fait référence est assez restreint, donc plus facilement maîtrisable ; les énoncés descriptifs sont relativement courts mais riches et extrêmement variés en terme de vocabulaire, de structures logiques, de mises en forme matérielles (Pouchot 1997).

Enfin, reproduire un corpus permet également de comparer les résultats obtenus (textes engendrés) aux textes naturels. Le processus est donc bouclé puisque le principe est de partir de textes écrits par des sujets, d' en extraire des modèles pour les intégrer au générateur afin qu' il produise des textes, destinés à être

lus, compris et éventuellement analysés par des humains (Denhière, Baudet 1992).

2. La génération : une aide à la recherche d'information

L'application précédemment décrite nous a permis de valider le générateur et d'y intégrer de nouvelles données pour en améliorer la production. Cette application a également mis en avant la nécessité de prendre en compte l'utilisateur. Profitant des travaux menés en parallèle au sein de notre équipe sur la Recherche d'Information (RI), nous étudions actuellement le passage à une autre application. Il s'agit ici de faciliter la recherche de l'utilisateur par l'association de notre générateur à un système de RI. La première partie de notre étude porte sur la problématique de l'interaction système de RI/utilisateur. La seconde s'intéresse aux apports de la génération de textes à la RI.

2.1 problématique de l'interaction homme-système en recherche d'information

Trois acteurs principaux interagissent lors d'une phase de recherche d'information : l'utilisateur, le Moteur de Recherche (MR) et la Base de Données (BD). Le MR et la BD forment le Système de Recherche d'Informations (SRI) auquel l'utilisateur accède via une interface. L'interaction entre ces trois acteurs se situe à différents niveaux. Pour répondre à un besoin réel d'information (RIN⁵) auquel correspond un ensemble (éventuellement vide) d'informations pertinentes (IP) dans la base de données, l'utilisateur interroge le SRI. Pour cela, il construit une représentation mentale (PIN⁶) de son RIN et l'exprime sous forme de requête à destination du SRI (1). Après le traitement de la requête par le MR en relation avec la BD (2), un ensemble de réponses (éventuellement vide), défini ici comme Résultat de la Requête (RR), est envoyé à l'utilisateur (3). La pertinence et l'exhaustivité de ces réponses dépendent de plusieurs facteurs ; elles sont de plus soumises à l'appréciation de l'utilisateur.

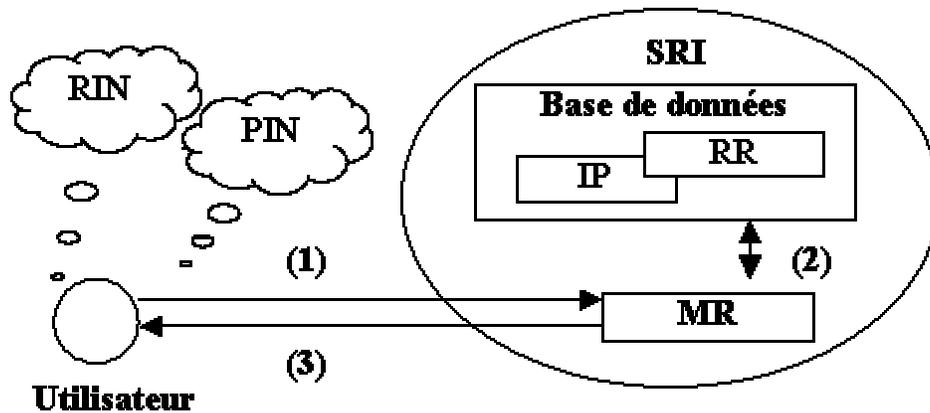


Fig. 2. Interactions en recherche d'information

L'un des problèmes de la recherche d'information consiste à obtenir le meilleur recouvrement entre l'ensemble des informations pertinentes et les résultats. Ce problème a deux origines.

- *Le passage du RIN au PIN.* La complexité de cette opération est essentiellement due au fait que l'utilisateur recherche des informations sur lesquelles il possède un état de connaissance justement incomplet. Il ne peut, ainsi, avoir qu'une perception incorrecte de son RIN. Ceci se reflétera sur son PIN et par transitivité sur sa requête.
- *Le passage du PIN à la requête.* Lors de la formulation de la requête, l'objectif est d'atténuer dans la réponse la production de "bruit" (réponses non pertinentes) et la production de "silence" (manque de réponses pertinentes). Afin de réduire la portée de ce double problème, la requête de l'utilisateur doit être adaptée aussi bien à la base de données qu'au moteur de recherche. A cette fin, seul l'utilisateur peut modifier sa requête initiale. Cependant, certains systèmes proposent des aides à l'utilisateur pour mener à bien cette modification. Par exemple, des systèmes comme AskJeeves⁷ proposent à l'utilisateur différentes reformulations de la requête initiale. D'autres comme DigOut4U⁸ utilisent un lexique de synonymes pour aider à reformuler la requête.

Les réponses du système constituent un deuxième écueil pour la recherche d'information. En effet, elles doivent être en adéquation avec la requête et

convenir à l' utilisateur. Si l' adéquation réponses/requête est un phénomène dû au moteur lui même, il n' en est pas de même pour l' adéquation réponses/utilisateur. Par exemple, un ensemble important de réponses pertinentes par rapport à la requête mais mal présenté risque de nuire à la compréhension de l' utilisateur ou, du moins, de lui compliquer la tâche. Un autre exemple est la production de réponses sous forme de listes alors que l' utilisateur attend un texte simple. Plusieurs systèmes prennent en compte la présentation des résultats et y apportent diverses solutions. Certains, comme Copernic⁹ ou BullEyes¹⁰, permettent différents tris des listes de résultats. D' autres, comme DigOut4U ou Inforian Quest¹¹, peuvent, à la demande, produire un court résumé des réponses.

Un troisième problème à résoudre est lié à l' appréciation des réponses par l' utilisateur. En effet, s' il est seul à pouvoir déterminer les réponses pertinentes parmi celles proposées, il ne possède, par contre, aucun critère lui permettant d' évaluer leur nombre. Dans le meilleur des cas, le système propose des réponses mais comment mesurer leur exhaustivité ? Dans le cas inverse, le système ne fournit aucune réponse :

- les informations pertinentes existent dans la base de données mais la requête de l'utilisateur n'est pas adaptée au MRI,
- aucune information pertinente n'existe dans la base de données. Ici l'utilisateur ne sait généralement pas à quoi est dû l'échec de son interrogation. En fait, il peut relever de différents facteurs : formalisation de la requête par l'utilisateur, fonctionnement interne du MRI, contenu de la base de données...

2.2 Intégration d' un générateur à un SRI

Nous considérons ici qu' un SRI est constitué de quatre éléments : un moteur de recherche, une base de données, un analyseur de requête et un générateur. Ces éléments sont partagés en deux catégories de traitement : traitement de l' information (moteur de recherche et base de données) et traitement linguistique (analyseur et générateur). Une interface fait le lien entre l' utilisateur

et le système de recherche d'information, et permet l'échange d'informations entre les deux.

Le processus d'interaction se décompose en deux étapes : de l'utilisateur vers le système et du système vers l'utilisateur. Nous présentons ici le rôle de notre générateur dans chacune de ces étapes.

2.2.1 De l'utilisateur vers le système : formulation de la requête

Comme nous l'avons déjà mentionné, une requête peut ne pas être adaptée au PIN et au moteur de recherche. Dans une telle situation, le système peut fonctionner de manière intelligente (transparence du processus de RI) ou encore de manière coopérative (interaction avec l'utilisateur), pour aider à la formulation et à la reformulation de la requête dans le but d'obtenir de meilleurs résultats. Le générateur, en produisant des paraphrases de la requête initiale, fournit à l'utilisateur toutes les verbalisations possibles de son besoin d'information. Ceci ne peut qu'avoir un impact sur la perception et la formulation de la demande par l'utilisateur.

Nous abordons alors la génération sous un angle un peu particulier et différent de l'approche classique. Le destinataire final du message généré est en fait un système et non pas un humain. Le rôle de l'utilisateur est alors de sélectionner les paraphrases à transmettre au moteur de recherche. Nous posons plusieurs hypothèses sur cette étape de paraphrasage.

La première porte sur le mode d'interrogation qui doit être en langue naturelle. En effet, si le moteur ne traite pas la langue naturelle, il est inutile générer automatiquement des requêtes sous forme textuelle pour les lui fournir.

La seconde hypothèse est que toutes les requêtes sont paraphrasées par le générateur même si elles sont compréhensibles et aboutissent à des résultats pertinents. Le générateur utilise des critères syntaxiques (correction syntaxique, modification ou simplification de la requête par exemple). Il fait aussi usage de quelques connaissances particulières sur le moteur de recherche telles que les mots clés de l'application. Les paraphrases doivent respecter le besoin exprimé par l'utilisateur. C'est pourquoi la méthode que nous adoptons pour préserver ce

besoin est de limiter le paraphrasage de la requête initiale à deux types de transformations : syntaxiques et lexicales (Ben Ali, Timimi 1999). De plus, le procédé de recherche d'information étant souvent basé sur l'appariement de termes, il semble pertinent de proposer plusieurs versions d'une même requête sans en modifier le sens afin d'augmenter le nombre de réponses potentielles.

Transformations syntaxiques

Cette reformulation consiste à transformer toute la structure syntaxique de la requête. Une requête peut être restructurée, par insertion ou effacement de mots vides de sens ou par permutation des termes, en une nouvelle requête qui a une syntaxe très différente. Cette transformation préserve ainsi les fonctions des constituants et le contenu informatif. C'est le cas des requêtes suivantes : “ *Augmentation de la consultation du Web ?* ” , “ *La consultation du Web est-elle en augmentation ?* ” , “ *La consultation du Web augmente-t-elle ?* ” .

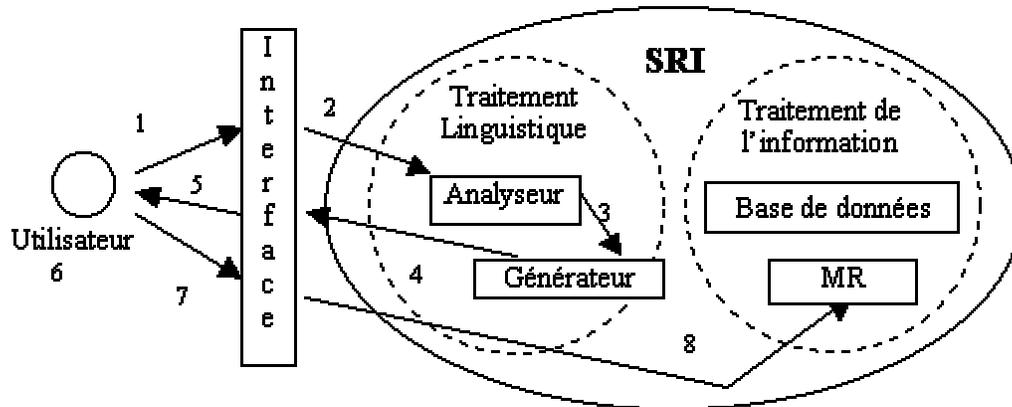
Notons également que le changement de voix de la requête (actif/passif) constitue une transformation souvent judicieuse.

Transformations lexicales

Les transformations paraphrastiques peuvent aussi être d'ordre lexical. Elles ne sont efficaces que si les mots utilisés sont recensés dans la base textuelle interrogée. Un fichier index par exemple peut contribuer à l'amélioration de la reformulation de requêtes. Il est alors possible d'employer les mots appropriés pour le paraphrasage des requêtes, particulièrement quand il s'agit de synonymie ou de mots à sens proche. Ceci nous confère une plus grande maîtrise des risques de modification du sens de la requête initiale. Les résultats obtenus suite à de telles transformations sont du type : “ *Augmentation de la consultation du Web ?* ” , “ *La consultation du Web est-elle en hausse ?* ” , “ *La fréquentation de la toile augmente-t-elle ?* ” .

Le générateur propose toutes les paraphrases possibles à l'utilisateur, qui choisit alors parmi ces possibilités celle(s) qui correspond(ent) le mieux à son besoin. Il peut sélectionner une ou plusieurs paraphrase(s) ainsi que sa requête initiale

pour les soumettre par la suite au traitement du moteur de recherche. Ces phases sont représentées dans le schéma ci-dessous :



Ce processus est composé de huit étapes :

1. L'utilisateur exprime son besoin d'information en langue naturelle puis l'envoie à l'interface du système ;
2. L'interface achemine la demande à l'analyseur ;
3. La requête est analysée et la représentation formelle de son contenu est fournie en entrée au générateur ;
4. Le générateur produit des paraphrases de la requête et les renvoie à l'interface ;
5. L'interface propose les paraphrases du générateur à l'utilisateur ;
6. L'utilisateur sélectionne une ou plusieurs paraphrase(s), incluant ou non sa requête initiale ;
7. La sélection est retournée à l'interface ;
8. L'interface soumet la sélection de requêtes au moteur de recherche.

2.2.2 Du système vers l' utilisateur : donner des résultats

Dans cette seconde étape, l' approche de la génération est plus classique. Le but est de transmettre des éléments d' information à l' utilisateur. Comme pour la première étape, nous considérons que tout le flux d' information passe par le

générateur avant d' être fourni à l' utilisateur final. Dans notre application, le générateur peut communiquer deux types d' informations.

Tout d' abord, il semble intéressant de fournir des éléments d' explication sur le fonctionnement du système de recherche d' information. Connaissant ce fonctionnement, l' utilisateur devrait être capable d' exprimer sa requête de façon mieux adaptée. L' objectif est que le système donne des réponses plus appropriées que “ pas de document trouvé ” . Il peut être possible d' expliquer à l' utilisateur quel est le domaine de la base, comment s' est effectuée la recherche, etc.

De plus, si le SRI contient par exemple un thesaurus, il est possible de fournir quelques indications d' extensions de recherche.

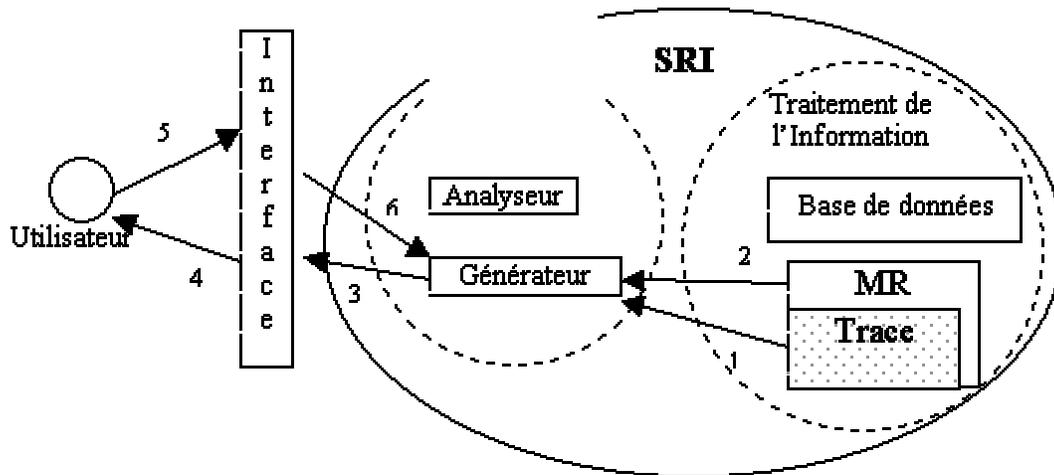
Le travail du générateur est donc basé sur une trace du processus de recherche qui constitue l' entrée de la production automatique en langue naturelle.

Le générateur peut également intervenir dans la présentation des résultats, fournis généralement sous forme de liste. Il a la possibilité d' améliorer la présentation en structurant, triant, ou même résumant les réponses du SRI.

L' utilisateur est alors en mesure de choisir l' affichage le plus approprié. Dans ce domaine, une étude sur la génération de commentaires de résultats bibliographiques a déjà été menée au sein de notre équipe (Ben Ali 1997).

Ces interactions sont représentées sur le schéma suivant :

Traitement Linguistique



Les six étapes présentées sont les suivantes :

1. La trace est donnée en entrée de la tâche de génération ;
2. Dans le même temps, les résultats de la requête sont également fournis au générateur ;
3. Après le traitement, le générateur transmet des données (exprimées en langue naturelle) à l'interface ;
4. L'interface envoie les résultats à l'utilisateur ;
5. L'utilisateur peut demander d'autres affichages à l'interface ;
6. L'interface interroge le générateur selon les demandes de l'utilisateur.

Les étapes 5, 6, 3 et 4 sont répétées tant que l'utilisateur souhaite des présentations différentes.

L'intégration du générateur au SRI permet ainsi de faciliter les formulations de textes, qu'ils proviennent de l'utilisateur ou du moteur de recherche. Dans le premier cas, le générateur assiste l'utilisateur dans l'expression de sa

requête ; dans le second, il améliore la compréhension de l'humain en lui fournissant des explications et des présentations de résultats plus adaptées.

3. Conclusion

Le générateur de notre équipe a été élaboré de manière à pouvoir traiter toute application. La première, brièvement décrite ici, nous a permis de tester cet outil et d'en améliorer les productions. Une collaboration entre les différents pôles de notre équipe (recherche d'information, génération automatique de textes et analyse) a été à l'origine de la deuxième application, plus complexe, qui est en cours de développement. Associé à un SRI, le générateur peut aider les utilisateurs dans leur démarche, en facilitant l'interaction personne-système durant un processus de recherche d'information. En effet, cette tâche peut s'avérer fastidieuse et complexe si le système interrogé fonctionne sur des techniques classiques (appariement, pondération, etc.), qui agissent uniquement au cœur du moteur de recherche. L'intérêt et les apports de l'utilisation de traitements linguistiques (génération et analyse) concernent ici deux aspects. L'interface est concernée en premier lieu. En effet, le recours à ces techniques augmente le degré de liberté de l'utilisateur en l'autorisant à verbaliser (formuler en langue naturelle) sa requête librement, c'est-à-dire sans contrainte de formalisation imposée par le système. Via une interface en langue naturelle un utilisateur peut exprimer au mieux son besoin en information.

Deuxièmement, si les techniques statistiques, par exemple, recherchent au niveau appariement une similarité parfaite entre plusieurs formes de surface (requête et contenu d'une base textuelle), les techniques linguistiques, quant à elles, autorisent en plus une ressemblance au niveau du contenu. Moyennant des phases d'analyse diverses (morphologique, syntaxique, sémantique et du discours) et appropriées à l'application, les outils linguistiques permettent de passer de la similarité de surface à la similarité des thèmes, des idées voire des concepts.

En ce sens, nous sommes persuadés de l'intérêt d'associer notre générateur à un SRI. En outre, concernant l'amélioration de la pertinence des recherches,

les différentes techniques (classiques et linguistiques) ne sont pas concurrentes mais complémentaires. L'évolution technologique des moteurs de recherche disponibles sur le marché ainsi que le nombre de partenariats actifs dans ce domaine montrent bien que si le problème n'est pas simple, la solution réside sans doute dans la combinaison des techniques disponibles (Leloup 1998). Nous souhaitons ainsi pouvoir matérialiser ces différentes pistes de recherche par la mise en œuvre d' une application réelle à partir d' un moteur de recherche existant.

Notes

¹ " Hourtin 99 : Lionel Jospin annonce une grande loi sur la société de l'information " , Le Monde, vendredi 10 septembre 1999.

² NUA, Internet Surveys, 31 mars 2000, <http://www.nua.ie/surveys/>

³ Institut National de Recherche en Informatique et Automatique

⁴ Passage d' une figure (deux dimensions) à un texte (une dimension)

⁵ Real Information Need (Mizzaro 1998).

⁶ Perceived Information Need (Mizzaro 1998).

⁷ Ce système (<http://www.askjeeves.com>) est un agent intelligent pour le Web (Revelli 2000)

⁸ Idem (<http://www.arisem.com>)

⁹ Ibidem (<http://www.copernic.com>)

¹⁰ Ibidem (<http://www.intelliseek.com>)

¹¹ Ibidem (<http://www.inforian.com>)

Bibliographie

Balicco L. (1993), *Génération de répliques en français dans une interface Homme-Machine en langue naturelle*, Thèse de doctorat, Université Grenoble 2, France.

Balicco L. (1997), " Génération automatique de descriptions textuelles de figures à deux dimensions " , GAT' 97, *1er colloque francophone sur la Génération Automatique de Textes*, Grenoble, France.

- Balocco L., Pouchot S. (1999), " Analyses d' un corpus en langue naturelle pour la génération automatique de textes descriptifs " , *Atelier Corpus et TAL, TALN' 99*, Cargèse, France.
- Ben Ali (1997), *Reformulation de réponses de systèmes documentaires bibliographiques*, mémoire de DEA, Université Grenoble 3, France.
- Ben Ali S., Timimi I. (1999), " De la paraphrase à la recherche d' information " , *CISI' 99*, ISD de Tunis, Tunisie.
- Bisseret A., Montarnal C. (1996), " Linearization in spatial descriptions: Tour or hierarchical structures? " *CPC*, 15 (5), pp. 487-512.
- Bourdoncle F. (1999), " Panorama et perspectives des outils de recherche d' Information textuelle sur Internet " , *Actes du colloque IDT' 99*, Paris, France.
- Dale R. (1989), *Generating referring expressions in a domain of objects and processes*, PhD Thesis, Edinburg, Ecosse.
- Danlos L. (1989), " Génération automatique de textes en langue naturelle " , *Ann. Télécommun.* 44, n° 1-2, France.
- Denhière G., Baudet S. (1992), *Lecture compréhension de texte et science cognitive*, Presses Universitaires de France.
- Fluhr C. (1992), " Le traitement du langage naturel dans la recherche d'information documentaire " in *Les interfaces intelligentes dans l'IST*, Rapport INRIA, France.
- Haton J.-P. et M.-C. (1993), *L' intelligence artificielle*, 3ème édition corrigée, Collection Que sais-je ?, Presses Universitaires de France.
- Horacek H. (1992), " An integrated view of text planning " , *Aspects of automated natural language generation, 6th International Workshop on Natural Language Generation*, Trento, Italie.
- Leloup C. (1998), *Moteur de recherche et d' indexation : environnements client-serveur, internet et intranet*, Paris : Editions Eyrolles.
- Miège B. (1989), *La société conquise par la communication*, Grenoble : Presses Universitaires de Grenoble.

- Mizzaro, S. (1998), How many relevances in information retrieval?, *Interacting with computers*, 10:3 305-322.
- Naït-Baha L., Jackiewicz A., Dijoua B. & Laublet P. (1999), “ Reformulation de requêtes pour la collecte d’ informations sur le Web à partir de points de vue ” , *Deuxième colloque du chapitre français de l’ ISKO*, Lyon, France.
- Nogier JF. (1991), *Génération automatique de langages et graphes conceptuels*, Paris : Editions Hermès.
- Ponton C. (1996), *Génération automatique de textes – Essai de définition d'un système noyau*, Thèse de doctorat, Université Grenoble 3, France.
- Pouchot S. (1997), *Etude de la structure de textes en langue naturelle en vue d’ une application en génération automatique de textes*, Mémoire de DEA, Université Grenoble 3, France.
- Revelli C. (2000), *Intelligence stratégique sur Internet*, Paris : Editions Dunod (à paraître).
- Sfez L. (1999), *La communication*, 5ème édition corrigée, Collection Que sais-je ?, Presses Universitaires de France