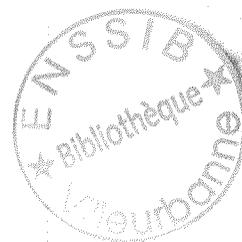


1084

M 1995 SYN DEA 01

enssib

école nationale supérieure des sciences de l'information et des bibliothèques



DEA en Sciences de l'Information et de la Communication

Option 3 : Systèmes d'Information Documentaire

Note de Synthèse

Les Systèmes de Recherche d'Information en Langage Naturel

Réalisée par **Hélio KURAMOTO**

sous la direction de **Mr Richard BOUCHE**

Villeurbanne - France

Avril 1995

BIBLIOTHEQUE DE L'ENSSIB



814312C

Table des Matières

Table des Illustrations	iii
INTRODUCTION	1
CHAPITRE I — Systèmes de Recherche d'Information (SRI)	4
1. Conception	4
2. Les SRI et les Systèmes de d'Information	4
3. Utilisation d'un SRI : problèmes de convivialité et de précision des résultats	6
CHAPITRE II — Systèmes de Recherche d'Information en Langage Naturel	11
1. Conception ou caractérisation	11
2. Typologies des SRI en Langage Naturel	12
2.1 Différents types d'interfaces	12
2.1.1 Interfaces orientés à arborescence de menus	13
2.1.2 Interfaces orientés au dialogue	14
2.1.3 Interfaces de navigation	14
2.1.4 Interfaces multimodales	17
2.2 Types de langage d'interface usager-système	18
2.2.1 Langages formels	18
2.2.2 Langages non formels	21
2.3 Traitement de la requête en langage naturel	23
2.3.1 Interfaces à traduction directe	24
2.3.2 Interfaces à traduction à un langage intermédiaire	25
2.3.3 Interfaces à modèle du domaine du discours avec des modèle utilisateur .	27
2.4 Représentation du contenu et évaluation des réponses	28
2.4.1 Représentation par mots ou groupe de mots	29
2.4.2 Représentation utilisant des base de connaissances	32
2.4.3 Représentation par structures syntaxiques	34
Conclusion	36
Références Bibliographiques	39
Bibliographie	44

Table des Illustrations

Tableau 1 - Recherche Documentaire	3
Figure 1 - Superposition entre quatre types de systèmes d'information	5
Figure 2 - Environnement d'un système d'information	6
Figure 3 - Schéma fonctionnel d'un système d'information	6
Figure 4 - Fonctionnement global du prototype de Mecabouche et Bassano	28

INTRODUCTION

Depuis que les ordinateurs sont apparus, des milliards d'informations y ont été enregistrées dans plusieurs bases de données, dans divers domaines de connaissances et sur diverses formes (numériques, textuels, imagés etc...). Etant donné que les ressources informationnelles sont de plus en plus accessibles aux utilisateurs personnels, le principal problème aujourd'hui est de savoir comment accéder à l'information dont on a besoin, de façon précise et conviviale. Pour cela, il faut donc utiliser des systèmes de recherche d'information (SRI).

Plusieurs types de SRI peuvent être trouvés dans le marché, selon les caractéristiques de l'information ou de l'application : les SQL (Structured Query Language) pour l'interrogation de bases de données relationnelles, les Systèmes de Recherche d'Information Documentaires pour les bases de données référentielles, les Systèmes Experts pour les bases de connaissances, les Systèmes Hypertexte pour les données organisées sous forme d'hyperdocuments, etc.

Plusieurs efforts ont été faits pour développer un SRI de manière à le rendre plus convivial et précis par rapport aux résultats de la recherche d'information. Parmi ceux-ci on trouve le Système de Recherche d'Information en *Langage Naturel*.

Langage naturel ou *langue naturelle*? La *langue*, selon le Robert, est un "système d'expression et de communication, commun à un groupe social (communauté linguistique) => idiome". Le terme *langue* est ainsi plus approprié à désigner des idiomes comme la *langue française*, la *langue anglaise*, la *langue portugaise* etc. Quant au terme *langage*, toujours

selon le Robert, est "la fonction d'expression de la pensée et de communication entre les hommes, mise en oeuvre au moyen d'un système de signes vocaux (parole) et éventuellement de signes graphiques (écriture) qui constitue une langue". Le dictionnaire Robert précise encore que, *langage* désigne tout systèmes de signes permettant la communication. Ainsi, on a le terme *langage machine* avec lequel on donne des instructions à un ordinateur. Le terme *naturel* est souvent trouvé dans le domaine du traitement automatique des langues, ceci pour préciser qu'il ne s'agit pas des langages formels ou des langages de programmation, etc... Encore, selon le Robert, "*langage naturel*, que représentent les langues du monde". Lorsqu'un langage utilise les termes et les structures de la langue on parle alors de *langage naturel* [POLITY].

Dans ce contexte, bien que les Systèmes Hypertexte correspondent plutôt à une procédure de navigation dans l'information qu'à un langage d'interrogation en vue de trouver l'information, ils sont tout de même à considérer du fait que des approches sémantiques pour la création des liaisons dans ces systèmes existent déjà.

Le but de ce travail est de présenter l'état de la littérature sur les systèmes de recherche d'information en langage naturel. On exposera dans un premier temps le fonctionnement de ces systèmes, ses structures et les problèmes de convivialité et de précision. On essaiera ensuite de définir SRI en Langage Naturel. Pour bien connaître les approches plus récentes, on fera une typologie des Systèmes de Recherche d'Information, suivi de la présentation de quelques prototypes et systèmes. On présentera dans la conclusion de ce travail quelles sont les caractéristiques la plus souhaitables pour un SRI en Langage Naturel et quelles sont les tendances actuelles. Les documents cités dans le corpus de ce texte

sont énumérés dans la liste Références Bibliographiques. Les autres documents non cités et/ou non consultés sont énumérés dans la liste Bibliographie.

La documentation nécessaire pour élaborer cette Note de Synthèse a été trouvée aux moyens de recherches réalisées sur les bases de données INSPEC, PASCAL et LISA. Le tableau 1 présente la stratégie utilisée et les résultats obtenus. Les résultats ont été limités aux dernières années en excluant ce qui concerne la conception et le fonctionnement d'un Système de Recherche d'Information.

Sources	Période	Stratégie de recherche	Résultat	Pertinence
INSPEC ¹	1993 - 1995	((information (w) retrieval (w) systems) and (natural (w) language*) not (speech (w) recognition)) / de,id	45	31
LISA ²	1990 - 1994	((su = natural language) or (su = computational linguistics)) and da > 90) and (su = information retrieval)	47	21
PASCAL ³	1992, 1994 ⁴	((li = interrogation et li = base* et li = donnee*) ou (li = recherche et li = information)) et (li = langage et li = naturel)	72	34
Autres ⁵				22
TOTAL			164	108

Tableau 1 - Recherche Documentaire

¹ Consultation réalisée au Centre de Documentation de l'ENSSIB et à l'URFIST

² Consultation réalisée sur CD-ROM à la bibliothèque de l'ENSSIB

³ Consultation réalisée sur CD-ROM à l'Université Claude Bernard - Bibliothèque Universitaire

⁴ Il n'a pas été possible de consulter l'année 1993 de la base PASCAL, la BU n'en a pas eu cette année

⁵ Documents obtenus au moyen de recherche documentaire dans la bibliothèque et dans le Centre de Documentation de l'ENSSIB

CHAPITRE I — Systèmes de Recherche d'Information (SRI)

1. Conception

Un Système de Recherche d'Information (SRI) est un dispositif que s'interpose entre les usagers potentiels et la collection d'information [HARTER]. De plus, il a une fonction typique, c'est la sélection des documents dans une base de données, en réponse à une requête de l'utilisateur, et leur rangement en ordre de pertinence [STRZALKOWSKI].

Selon Smeaton, ce qu'un SRI propose est la recherche de documents en réponse à une requête de l'utilisateur de manière à ce que le contenu de ces documents corresponde au besoin originel d'information de l'utilisateur [SMEATON].

Salton en conclusion affirme qu'un SRI traite de la représentation, du stockage, de l'organisation et de l'accès aux items d'information [SALTON et MCGILL].

On peut dire alors qu'un SRI est d'une part une interface (pour interagir avec les usagers), des mécanismes afin de déterminer la pertinence ou non des documents par rapport à la demande des usagers et d'autre part une structure de données organisée de manière à permettre l'accès rapide et précis à l'information.

2. Les SRI et les Systèmes d'Information

Les Systèmes de Recherche de l'Information présente des points de similitude avec plusieurs domaines de traitement d'information. Les plus importants systèmes d'information

automatisés sont les systèmes de gestion d'information, les systèmes de gestion de bases de données, les systèmes de support à la décision, les systèmes de question-réponse, ainsi que les systèmes de recherche d'information. La figure 1 montre les relations entre les plusieurs types de systèmes d'information [SALTON et MCGIL].

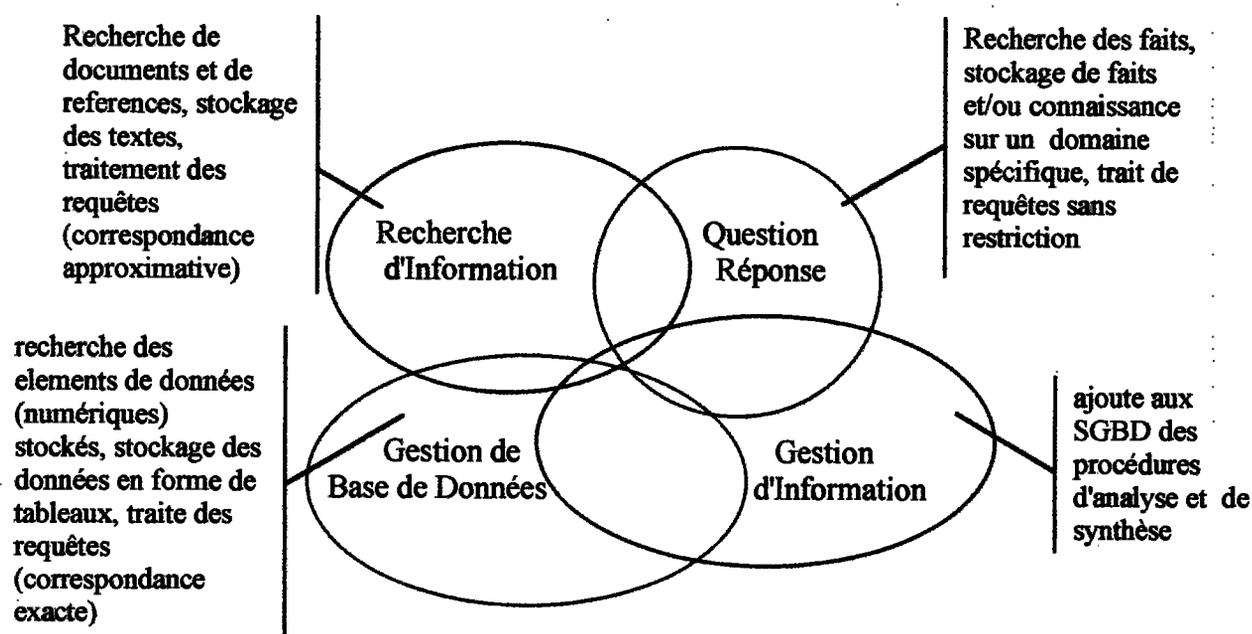


Figure 1 - Superposition entre les quatre types de systèmes d'information

Les évolutions récentes des technologies de l'information ont modifié ce paysage et multiplié les banques de données en texte intégral, les applications de la gestion électronique de documents et surtout les systèmes d'information multimédias [POLITY].

Un point en commun entre tous ces systèmes d'information est l'interface nécessaire pour que l'utilisateur puisse avoir accès aux informations.

Tout SRI peut être décrit comme un système constitué d'un ensemble d'information (DOCS), un ensemble de requêtes (REQs) et d'un mécanisme (SIMILAIRE) qui détermine quelles informations répondent aux requêtes [SALTON et MCGIL].

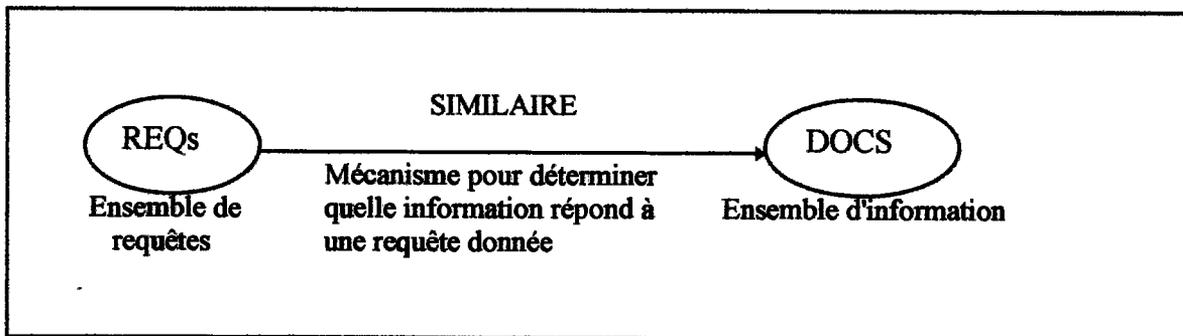


Figure 2 - Environnement d'un système d'information

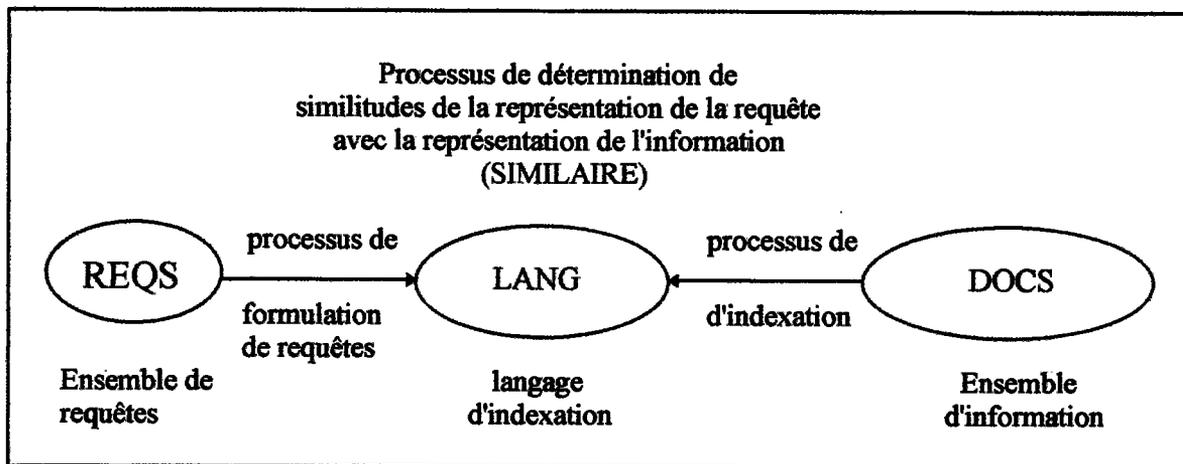


Figure 3 - Schéma fonctionnel d'un système d'information

3. Utilisation d'un SRI : problèmes de convivialité et de précision des résultats

Face à la prolifération des sources d'informations et à l'utilisation généralisée d'outils informatiques de plus en plus évolués, l'utilisateur final veut disposer de systèmes d'utilisation simples et conviviaux lui donnant rapidement une vision synthétique des informations, mais aussi des réponses pertinentes à ses questions précises [CORET et al.].

Autrement dit, les systèmes d'information traditionnels possèdent des interfaces dits *à la commande* qui exigent de l'utilisateur la maîtrise d'un grand nombre de connaissances de caractère hétérogène puisque certaines sont d'ordre informatique, d'autres concernent la

structuration des données, et d'autres enfin ont trait au vocabulaire du domaine. Ces systèmes sont des systèmes professionnels et demandent donc, comme pour tout outil professionnel, une certaine formation. On listera ici quelques éléments dont la maîtrise conditionne l'utilisation de ces systèmes [POLITY] :

- ◊ un certain nombre de commandes pour se mettre en mode interrogation, pour écrire une requête dans une syntaxe correcte, pour visualiser les résultats, pour les imprimer, etc.;
- ◊ la logique booléenne indispensable dans une recherche multicritères. Intersection, union, exclusion sont les prérequis pour l'exécution d'une bonne stratégie de recherche ;
- ◊ la troncature et les différents opérateurs de comparaison, de proximité etc... ;
- ◊ la structure conceptuelle de la base de données, les noms des champs à interroger et les conventions d'écriture dans chacun de ces champs ;
- ◊ les termes d'indexation, les lexiques, les thesaurus, etc...

La convivialité des interfaces de SRI a été mise sous épreuve avec la croissance soudaine des usagers de bases de données dans divers secteurs d'activités. Pour avoir accès à l'information, l'utilisateur doit se familiariser avec le SRI et dominer quelque langage d'interrogation de base de données. Cette difficulté de familiarisation rend difficile la croissance du nombre de ces usagers [SHOW et al.].

Le plus grand défaut des interfaces de SRI ou langages de recherche est qu'elles sont de manière générale difficiles à apprendre et à utiliser, spécialement pour les usagers non spécialisés. Cela est dû au fait qu'elles ont des règles rigides de grammaire et des expressions mathématiques compliquées (dans le cas du SQL) [WANG].

En plus des problèmes de convivialité, la recherche d'information traditionnelle trouve des problèmes de précision quant aux résultats de la recherche.

Souvent exprimés sous forme d'une combinaison booléenne de mots clés, une recherche d'information peut s'exprimer à l'aide d'une spécification de proximité ou même inclure les poids (déterminés automatiquement ou par définition de l'utilisateur) sur les mots clés. Cependant, ce type de recherche peut être non efficace, principalement pour les raisons suivantes :

- ◊ les mots du texte peuvent avoir des sens ou des significés différents selon le domaine. Exemple: *Sharp* peut indiquer une intensité de douleur en médecine ou une mesure de qualité d'un outil de coupe en travaux de bois ;
- ◊ les mêmes mots peuvent être utilisés en différentes phrases, tout en ayant des liaisons différentes et exprimer des concepts totalement dissemblables. Les *Juvenile victims of crime* and *victims of juvenile crime* utilisent les mêmes quatre mots mais ont un autre ordre et signifié différent ;
- ◊ des mots complètement différents peuvent être utilisés pour exprimer le même concept. Par exemple, les phrases *Stomach pain after eating* and *prandial*

abdominal discomfort ont le même sens mais les mots utilisés dans la phrase sont complètement distincts.

Les approches conventionnelles utilisées en Recherche d'Information telles que l'indexation basée sur les mots et la recherche par expressions booléennes ne peuvent pour autant pas résoudre ce type de problème [SMEATON²].

L'évaluation d'une requête dans la recherche d'information peut être vue comme une comparaison entre les documents et la requête (ou entre leurs représentations). Dans le modèle classique [SALTON et MCGILL], cette comparaison est faite directement. C'est-à-dire, une comparaison basée sur l'utilisation commune des mots clés entre le document et la requête. Le point critique de ce modèle, est que si un document sémantiquement pertinent n'est pas représenté par les mêmes mots clés de la requête, alors le document sera jugé non pertinent pour le système. Par exemple, en un SRI où la recherche est faite aux moyens d'expressions booléennes, un document sur « document retrieval » ne sera pas trouvé si on utilise une requête telle que « information retrieval » [NIE].

Par conséquent, plusieurs chercheurs travaillent sur les SRI de manière à améliorer la précision des résultats et à faciliter l'accès à l'information avec un minimum de coût de familiarisation [SHOW et al.].

Ainsi, dans le but de rendre le Système de Recherche d'Information plus convivial et pour garantir des résultats plus précis, plusieurs chercheurs travaillent sur le développement et l'amélioration des SRI en langage naturel.

Le développement des interfaces d'un système d'information en langage naturel est souhaitable du point de vue de l'utilisateur car il n'a pas besoin d'apprendre un nouveau langage de commande artificielle de même qu'est plus facile pour l'utilisateur d'exprimer ses besoins d'information en langage naturel [SMEATON¹].

CHAPITRE II — Systèmes de Recherche d'Information en Langage Naturel

1. Conception ou caractérisation

On trouve dans la littérature spécialisée plusieurs références traitant des Systèmes de Recherche d'Information en Langage Naturel. Cependant, il est difficile d'établir une définition précise d'un SRI en Langage Naturel car il y a une variété de systèmes, dits en langage naturel, mais que présentent différentes manières de communiquer avec les usagers. Il faut donc mieux établir une typologie de ces systèmes que les définir.

La caractéristique la plus commune à ces systèmes dans la littérature spécialisée, est que ces usagers peuvent faire leurs demandes de recherche sous forme de phrases, en anglais, en français, en espagnol, etc... [ROWE et GUGLIELMO] [DAMI et LALLICH-BOIDIN] [SHOW et al.].

Une autre caractéristique possible des SRI en Langage Naturel est l'utilisation des structures syntaxiques. Il y a alors comparaison entre les structures syntaxiques des documents et de la requête [SMEATON et SHERIDAN].

D'autre part, il y a d'autres systèmes, comme par exemple WAIS, où les requêtes sont formulées en langage quasi-naturel, étant donné que celles-ci sont formulées au moyen de mots ou de phrases [KAHLE et al.].

Si on considère que tous les systèmes qui manipulent des requêtes constitués soit par de phrases avec une structure syntaxique, soit par de groupes de mots combinés (dépourvues de structure syntaxiques) ou même de mots de la langue, sont des SRI en Langage Naturel, il faudra alors inclure dans ceux-ci les systèmes traditionnels de recherche bibliographique

d'information. En effet ils manipulent des requêtes constituées d'expressions booléennes qui sont une combinaison des mots et opérateurs (*et, ou, sauf*), eux aussi de mots de la langue.

Ainsi, le critère pour définir un système comme étant en langage naturel ou non, ne peut pas se restreindre à la seule façon dont la requête est formulée. Alors il faut faire une typologie pour mieux les caractériser.

2 Typologie des SRI en langage naturel

A partir de la littérature trouvée sur ces systèmes on peut établir une typologie selon les critères suivants : 1) les différents types d'interfaces ; 2) les types de langage du dialogue usager-système ; 3) le traitement de la requête ; 4) la représentation du contenu et l'évaluation de la réponse.

2.1 Différents types d'interfaces

Plusieurs systèmes d'interaction entre l'utilisateur et l'ordinateur pour l'accès à l'information sont connus et peuvent être classés en quatre types : 1) interfaces orientés à arborescence de menus ; 2) interfaces orientés au dialogue ; 3) interfaces de navigation ; 4) interfaces multimodales. Eventuellement, il est possible de trouver des interfaces qui combinent les avantages des types cités ci-dessus avec d'autres de manière à améliorer la convivialité.

2.1.1 Interfaces orientés à arborescence de menus

Ces interfaces ont l'avantage de la simplicité puisque l'utilisateur doit seulement indiquer, à chaque étape, l'option choisie ; ces systèmes efficaces lorsqu'ils sont bien adaptés au public pour lequel ils sont préparés, ont toutefois l'inconvénient d'être très rigides : d'abord, comme on ne peut pas proposer un choix trop important, ils ne peuvent être conçus que pour un type d'utilisateurs relativement bien défini ; en suite, le déroulement de ces menus devient vite insupportable à l'utilisateur expérimenté (or une personne, novice lors de la première interrogation, s'exerce rapidement si elle pose plusieurs questions d'affilée) [GUIMIER-SORBETS].

Les défauts des interfaces à menus ont été à maintes reprises soulignés : rigidité, lenteur du déroulement, inefficacité de la recherche lorsque elle porte sur plusieurs critères (seul l'opérateur « et » est généralement disponible). Cependant, on constate que pour une utilisation occasionnelle, et pour des applications très ciblées ne mettant pas en jeu un vocabulaire très étendu, elles ont un grand succès. Ces interfaces sont devenues courantes à l'heure actuelle [POLITY]. Exemples : les applications bancaires ouvertes au grand public, les applications touristiques etc. Pour les applications de recherche d'information documentaire il y a un système que s'appelle CANSEARCH.

CANSEARCH est un système de recherche d'information sur la littérature concernant la thérapie du cancer dans la base Medline. Il a été développé par [POLLITT], ayant comme but les utilisateurs novices. Il est basé sur une arborescence de menus où les utilisateurs font la sélection des options en touchant un écran tactile. Les utilisateurs n'utilisent pas un clavier. Les utilisateurs construisent la requête en sélectionnant les mots ou phrases présentés dans les différents menus sur l'écran.



Ce type d'approche est connu comme Menu en Langage Naturel et il utilise des bases de connaissances pour la représentation du contenu des documents. Thompson et ses collègues ont développé un ensemble d'outils pour la construction d'interfaces de recherche d'information orienté à menus en langage naturel [ALBERICO et MICCO].

2.1.2 Interfaces orientés au dialogue

Ces interfaces ont soit un langage formel soit un langage non formel plus proche du langage courant des hommes, pour faire l'interaction entre les systèmes et les utilisateurs. Un exemple d'interface à langage formel (voir la section 2.2.1, Langages Formels) est celle utilisée pour les Systèmes de Recherche Documentaire (STAIRS, BRS, etc...) et celle utilisée pour les bases de données relationnelles que s'appellent SQL - *Structured Query Language* (ORACLE, INGRES, etc...).

Les récents développements des techniques du Traitement Automatique du Langage Naturel, les outils de l'Intelligence Artificielle, et l'aide des outils statistiques et des probabilités, ont permis la conception d'autres interfaces utilisant des langages plus proches des ceux couramment parlés pour les humains (voir le paragraphe 2.2.2, Langages Non Formels). Comme exemple on peut distinguer : le Spirit, le SCISOR, le WIN, le PBS, etc...

2.1.3 Interfaces de navigation

Ces interfaces comprennent les systèmes hypertextes et hypermédias. Le mot hypertexte a été créé par Theodore H. Nelson pour désigner des écrits non-linéaires. Les

caractéristiques principales des hypertextes sont celles qui définissent les hyperdocuments [BALPE] :

- ◊ ils sont constitués d'un ensemble d'informations enregistrées, uniquement de nature textuelle, au sein desquelles les parcours de lecture peuvent être multiples ;
- ◊ ils utilisent le multifenêtrage : plusieurs documents lus pouvant s'inscrire simultanément sur un même espace-écran ;
- ◊ ils permettent une certaine interactivité : la possibilité offerte au lecteur de faire ses choix de parcours localement, c'est-à-dire, à la différence des parcours de bases de données par exemple, non pas au départ de la demande qui est alors traitée globalement par l'ordinateur, mais en fonction de la lecture qu'il est en train de faire.

Les systèmes hypermédia sont, en effet, un ensemble d'informations appartenant à plusieurs types de médias (textes, son, image, logiciels) pouvant être lus, écoutés ou vus suivant les multiples parcours de lectures, en utilisant également la possibilité de multifenêtrage. Ce que différencie essentiellement l'hypermédia de l'hypertexte, n'est ainsi que la nature symbolique des codages d'information utilisées. Un hypermédia n'est rien d'autre qu'un hypertexte gérant des textes supportés par des médias divers [BALPE]. Exemples des systèmes hypertexte / hypermédia : GUIDE, HYPERCARD, SUPERCARD, HYPERDOC, WWW etc...

L'utilisation du traitement linguistique dans la documentation pour les développements des systèmes de recherche d'information en langage naturel n'est pas en concurrence avec les approches de type hypertexte, dans la mesure où la recherche en accès

direct et par la navigation sont complémentaires. De plus les outils de calcul de proximité sémantique entre textes peuvent devenir les moyens d'aide à la construction de liens hypertextes soit lors de l'élaboration des hypertextes soit dynamiquement lors de l'utilisation, en fonction du besoin exprimé par l'utilisateur [FLUHR].

Il existe actuellement quelques travaux utilisant des outils de traitement automatique du langage naturel pour améliorer ce type de système. Comme exemples de produits on peut distinguer : les logiciels CoDHIR [MAREGA et PAZIENZA], ConQuest [ADDISON et NELSON] et VISAR [CLITHEROW et al.].

Exemples :

a) CoDHIR [MAREGA et PAZIENZA]

CoDHIR est un système de recherche d'information orienté à navigation hypertextuelle. Il accomplit d'une part les fonctions spécifiques d'un système de recherche d'information (l'évaluation de pertinence parmi les documents et entre les requêtes et les documents, ainsi que les documents trouvés) et d'autre part les fonctions hypertextuelles (liaisons sémantiques parmi les documents, modules pour la navigation dans une base de données bibliographique, etc...) en utilisant des méthodologies basées sur les outils de traitement automatique du langage naturel.

b) ConQuest [ADDISON et NELSON]

ConQuest est un système de recherche d'information textuelle en langage naturel, basé sur concepts et commercialisé par ConQuest Software Inc. Il possède des fonctions de recherche et de butinage qui lui permet de automatiquement trouver des concepts liés dans un document en utilisant du traitement automatique du langage naturel et de la construction d'une base de connaissance. Il peut lier les

concepts des documents dynamiquement en agissant comme un hypertexte intelligent. Les techniques utilisés incluent : traitement automatique du langage naturel ; méthode de probabilité pour la détermination de pertinence des documents ; méthode pour lier statistiquement les documents en considérant la présence de concepts ; réseaux sémantiques et graphes conceptuels.

2.1.4 Interfaces multimodales

Actuellement, avec le développement de la technologie permettant d'utiliser des outils graphiques et des accessoires comme la souris, quelques interfaces ont été construites avec des différents modes d'expression. Le terme « mode » utilisé dans ce contexte, signifie le moyen d'exprimer. Il suppose l'existence d'un « médium » correspondant pour transporter l'information. En effet, la notion d'interface multimodale n'a de sens que dans un contexte de station multimédia. L'idée est que l'opérateur soit capable de dialoguer avec le système utilisant le mode le plus adéquat à son niveau de connaissance. Quelques exemples de modes disponibles : le graphique (utilisé pour dessiner plans, diagrammes etc...), le gestuel (utilisé pour faire mouvementer des objets, ex. : la souris), en langage naturel (utilisé pour décrire une situation, une requête etc...), en langage à la commande (utilisé pour dialoguer avec les systèmes) [DAMI et LALLICH-BOIDIN].

Exemple :

ESPRIT II/MMI2 - Multi-Mode Interface for M-Machine Interaction

C'est un système expert et a comme but la construction d'un système de communication homme-machine capable de travailler en plusieurs modes d'expression (langage naturel, graphique, gestuel et langage à la commande). L'idée

est de développer une interface interactive dont un opérateur humain puisse dialoguer avec un système expert. Ainsi, il est très important que le projet soit conçu en tenant compte d'abord de la conduite de l'utilisateur. Pour cela, font partie de l'équipe du projet des psychologues, et des informaticiens. Le système consiste en plusieurs modules « experts ». Tous les modules travaillent avec une unique base de communication représentée par une expression que s'appelle CMR (*Common formalism for representation of meaning*). Une expression CMR consiste en plusieurs parties : une logique de proposition de première ordre ; des opérations d'énonciation ; des informations syntaxiques et chronologiques ; des « *illocutionary force* ». Sur cet aspect, le CMR est une commande de description syntaxique. Ainsi le système transforme les commandes en langage naturel en expressions CMR, et vice versa. Pour cela, le système réalise l'analyse linguistique en trois stages : 1) analyse morphologique ; 2) analyse syntaxique (son fondement linguistique et son application sous forme d'un système expert) ; 3) expressions logiques [DAMI et LALLICH-BOIDIN].

2.2 Types de langage d'interface usager-système

Essentiellement on peut trouver dans cette typologie deux classes de langages : 1) langages formels ; et 2) langages non formels.

2.2.1 Langages formels

Les principales caractéristiques de ce type de langage viennent du fait qu'il est artificiellement défini, qu'il possède des règles rigides de grammaires et qu'il utilise des expressions complexes (booléennes et/ou mathématiques) [WANG].

Par conséquent, l'utilisateur non expert trouve des difficultés à les utiliser. On trouve d'autres applications à ce type de langage principalement celles de la recherche documentaire (langage à la commande basé sur des expressions booléennes) et les interfaces SQL pour la recherche sur les bases de données relationnelles.

En ce que concerne les applications de la recherche documentaire ou la recherche de texte intégral, les problèmes d'utilisation, de convivialité et de précision ont été discuté dans le chapitre précédent (voir Chapitre I, section 3).

Exemples d'utilisation de ce type de langage sont les grands serveurs de bases de données (DIALOG, STN, QUESTEL, etc...), que disposent des logiciels capables d'interroger ces bases permettant opérations comme [DEWEZE] :

- ◊ formulation des notions recherchées à l'aide d'opérandes que sont les mots-clés en langage naturel ou les descripteurs (unitermes ou syntagmes) en langage contrôlé (éventuellement à partir d'un lexique visualisable) ;
- ◊ combinaison booléenne des opérandes à l'aide d'opérateurs logiques (ET, OU, SAUF) ;
- ◊ combinaison éventuelle des opérandes avec des opérateurs de distance (mots adjacents dans un ordre donné, mots présents dans la même phrase - ou paragraphe - d'une référence) ;

Exemple : **information (w) retrieval (w) systems** => cette expression indique qu'on cherche les références que possèdent les mots *information*, *retrieval* et *systems* disposé adjacemment (l'opérateur (w) indique que les mots doivent être

adjacents) comme : *information retrieval systems*. On améliore la précision du résultat.

- ◊ recherche d'une chaîne de caractères dans le texte (titre et résumé), par exemple pour la recherche de composés chimiques, éventuellement à l'aide de troncatures ou de masques (à gauche ou à droite) remplaçant des préfixes ou d'autres chaînes de caractères ;

Exemple : **biblio+** => cette expression indique qu'on cherche tous les mots avec la base « biblio », comme résultat on peut obtenir des références avec les mots : **bibliobus, bibliographie, bibliographique, bibliométrie, bibliophile, bibliophiles, bibliophilie, bibliothéconomie, bibliothèque.**

Il est évident que le temps de réponse de toute recherche documentaire utilisant les possibilités d'interrogation au moyen de descripteurs tronqués à droite sera directement proportionnel au nombre de descripteurs du lexique répondant au critère de troncature. En conséquence, l'opérateur de troncature ne doit pas systématiquement être utilisé pour l'interrogation d'une base de données ; ne jamais oublier que la consultation du lexique est aussi une façon sûre de lever une ambiguïté. Les temps de réponse les meilleurs et la pertinence parfaite des réponses sont obtenus à partir des descripteurs non tronqués.

- ◊ visualisation des résultats en ligne ;
- ◊ impression des résultats en différé ;
- ◊ sauvegarde du questionnaire et possibilité de l'exécuter ultérieurement.

Pour ce type de système on constate l'existence, encore rudimentaire, d'un traitement linguistique du contenu des documents, ce qui permet la troncature et la recherche en utilisant des mots.

Par rapport aux interfaces du type SQL, au-delà de la complexité du langage, le plus grand problème est qu'elles ne sont pas capables d'exprimer des concepts vagues, étant donné qu'elles n'expriment que des valeurs exacts ou des intervalles de valeurs. Ainsi, pour une application factuelle comme les systèmes d'information géographiques, où on peut trouver souvent des concepts vagues comme « très chaud », « haute élévation », « bas température », etc..., ces interfaces ne sont pas simples à utiliser. Il faut bien connaître les opérateurs mathématiques et la manière de les placer selon l'ordre logique défini pour la langue [WANG]. Exemple de requête SQL pour trouver, dans une base de données d'informations géographiques, jusqu'à 5 villes qui sont au moins de 25 Km des campings et aussi qui sont au moins de 6,25 Km des centres historiques :

```
SELECT name FROM city
WHERE 5 ≤ (SELECT count (camp_name)
FROM distance_to_camping_site
WHERE city_name=CITY.name and distance ≤ 25) and
5 ≤ (SELECT count(hist_name)
FROM distance_to_historic_spot
WHERE city_name=CITY.name and distance ≤ 6,25)
```

2.2.2 Langages non formels

La caractéristique principale de ce type de langage est qu'il est plus proche de la langue parlée, sans la complexité d'un langage formel. La construction de ces interfaces est basée sur les techniques du traitement automatique du langage naturel pour la représentation du contenu des documents et des requêtes. Pour la détermination de la similarité entre les

documents et les requêtes, d'autres techniques sont aussi utilisées comme les outils statistiques, de probabilité, de systèmes experts, etc.

Les principaux problèmes trouvés dans ce type d'interface sont : la synonymie (plusieurs mots avec le même sens), la polysémie (mots à plusieurs sens ou significés [MAULDIN]), l'anaphore (phénomène où un mot, de la classe des pronoms, fait référence à un mot antérieur dans une phrase [DAMI et LALLICH-BOIDIN]), l'ellipse (omission d'un ou de plusieurs mots dans une phrase et qui rend cependant celle-ci compréhensible). De plus, il y a le problème de la dépendance du domaine, qui rend l'utilisation de ces interfaces presque toujours restreinte à un seul domaine.

Exemples des interfaces avec langages non formels:

a) PBS - Parsing, Boolean Recognition, Stemming

C'est une interface en langage naturel pour un système de recherche d'information bibliographique. Les requêtes sont soumises en langage naturel. L'interface PBS essaye d'identifier les termes et les opérateurs booléens implicites dans la requête. Cette interface permet la reformulation de la requête utilisant un morceau d'un document trouvé comme résultat d'une recherche antérieure et indiqué par l'utilisateur comme étant celui le plus représentatif de son besoin d'information. [SAMSTAG-SCHNOCK et MEADOW]

b) ESPRIT SIMPR

Ce projet a comme but de traiter les documents techniques à texte intégral de manière à extraire les termes qui puissent représenter précisément ses contenus. Un autre but est l'investigation des techniques pour la classification automatique de documents, l'apprentissage de la machine dans la classification de documents et la

recherche de documents techniques. L'approche utilisée pour la recherche d'information est basée sur la comparaison des phrases de la requête et des documents de la base de données. Le traitement automatique du langage naturel est réalisé aux moyens des deux analyses indépendantes : l'analyse morphologique et l'analyse syntaxique. L'analyse syntaxique est faite selon la grammaire CGP (*Constraint Grammar Parser*). L'avantage de cette approche est qu'elle est efficace, indépendante du domaine et qu'elle a besoin d'un dictionnaire mais non d'une base de connaissances comme ce qui est demandé pour l'analyseur sémantique. Par contre, elle génère et définit la structure du texte et non le contenu ou la valeur sémantique. Elle ne peut pas résoudre non plus les plusieurs types d'ambiguïtés.

Pour la représentation du contenu des documents et des requêtes elle génère, à partir de l'analyse linguistique, une structure d'Arbre Structuré Analytique (*Tree Structured Analytics - TSA*).

Pour la détermination de la similarité entre les documents et la requête ce système adopte une comparaison à trois niveaux: comparaison de mots, de la structure TSA elle-même et de la structure résiduelle. Les problèmes d'ambiguïté sont traités au niveau de l'analyse morpho-syntaxique, bien que d'autres ambiguïtés existent elles ne sont pas traitées ici [SMEATON et SHERIDAN].

2.3 Traitement de la requête en langage naturel

Cette typologie ne concerne que les interfaces dont le dialogue est en langage naturel.

Selon Binot, on peut distinguer trois grandes générations d'interfaces de dialogue : a)

interfaces à traduction directe ; b) interfaces à traduction à une langage intermédiaire ; c) interfaces à modèle du discours avec des modèle d'utilisateur [BINOT et al.²].

2.3.1 Interfaces à traduction directe

Les premières interfaces développées étaient « à traduction directe », c'est-à-dire qu'elles se composaient essentiellement d'un analyseur-traducteur transformant directement l'énoncé du langage naturel dans un formalisme adéquat pour l'application visée (par exemple un formalisme d'interrogation de bases de données) [BINOT et al.²].

Exemple:

- a) Un prototype d'interface, en langage naturel, pour les systèmes d'information géographique. [WANG]

Plusieurs concepts liés aux données géographiques sont vagues alors que les interfaces d'interrogation de bases de données relationnelles ont été développées pour n'exprimer que les concepts précis. Par exemple, un usager veut savoir d'un Système d'Information Géographique quelles sont les villes qui ont une température moyenne, une très basse humidité et une haute altitude. Or, les interfaces traditionnelles ne sont pas capables d'exprimer les conditions de sélection comme « moyenne », « basse », « haute » et le modificateur « très ». Ils peuvent néanmoins spécifier les valeurs précises ou les intervalles de valeurs. Ainsi, Fangju Wang propose un prototype d'interface en langage naturel ajouté à une approche basée sur les concepts d'ensemble flou pour résoudre les problèmes d'expression des concepts vagues (comme moyenne, très basse, haute, peu etc.). Ce prototype utilise des outils linguistiques pour traiter les requêtes et les traduire par une requête SQL.

b) IDQS - Intelligent Database Query System [SHOW et al.]

IDQS est une interface intelligente pour les bases de données relationnelles. Elle est constituée de deux grands ensembles d'outils: l'unité de traitement automatique du langage naturel (*Left-Associative Grammar* [HAUSSER], [KONG et SHOW]) et la base de connaissances. Elle suit le principe de traduction des requêtes en langue anglaise pour le SQL.

2.3.2 Interfaces à traduction à un langage intermédiaire

La deuxième génération d'interfaces est basée sur la notion de « langage intermédiaire ». Ce second modèle, adopté par la grande majorité de systèmes actuels, divise le processus de compréhension en deux étapes [BINOT et al.²] :

- a) un analyseur général traite l'énoncé en langage naturel pour produire une représentation de son contenu sémantique dans un « formalisme intermédiaire de représentation du sens » ;
- b) un interprète, dépendant de l'application visée, examine le contenu de la représentation intermédiaire et décide des opérations qu'il importe d'effectuer dans le contexte de l'application, ce qui comprend normalement la traduction de tout ou de partie de la forme intermédiaire dans un formalisme spécifique à l'application.

Exemple :

- a) LOQUI [BINOT et al.²]

LOQUI est une interface pour les bases de données relationnelles, basée sur l'usage d'un formalisme intermédiaire et d'un modèle conceptuel explicite du domaine du discours. Les principaux éléments de cette interface sont :

- ◊ traitement lexicale et morphologique de l'énoncé, permettant d'identifier un "lexique dynamique" (liste de sens possibles de tous les mots) ;
- ◊ un analyseur génère à partir de ce lexique dynamique une représentation sémantique interne représentant le sens de la phrase ;
- ◊ cette représentation sémantique intermédiaire fait l'objet d'un processus d'interprétation sémantique et pragmatique dont un des aspects réside dans l'extraction des informations appropriées de la base de données ;
- ◊ un processus de détermination de la réponse génère, à partir du résultat fourni par la base de données et de diverses informations relatives à l'énoncé et au contexte du dialogue, une nouvelle forme sémantique intermédiaire décrivant le contenu de la réponse à générer ;
- ◊ un générateur "tactique" produit une réponse en langage naturel à partir de la forme sémantique qui lui est fourni ;

- ◊ les processus qui précèdent s'appuient sur un modèle conceptuel du domaine d'application et sur un modèle explicite de la structure du dialogue en cours et des informations liées au contexte de ce dialogue [BINOT et al.²].

2.3.3 Interface à modèle du domaine du discours avec des modèle utilisateur

Enfin on peut distinguer une troisième génération d'interfaces, caractérisée par le fait qu'elles incluent, outre un modèle explicite du domaine du discours, un modèle explicite de l'utilisateur [BINOT et al.²].

Exemples :

a) ESPRIT / MMI2 [BINOT et al.¹]

b) Prototype (une extension de DIALECT [BASSANO])

Prototype pour la recherche documentaire basé sur les outils de traitement automatique du langage naturel et des systèmes experts. Il part d'un modèle linguistique réalisé par plusieurs experts qui coopèrent entre eux pour :

- a) extraire des documents proposés par un système documentaire, l'essentiel de l'information intéressante pour une application de type documentaire ;
- b) construire un réseau de lexis permettant ainsi de relancer automatiquement la recherche documentaire. Il introduit, aussi, une méthode de reformulation qui permet, à partir des documents pertinents, de regrouper les mots qui peuvent être en relation, dans une classe sémantique sans utiliser de thesaurus [MEKABOUCHE et BASSANO].

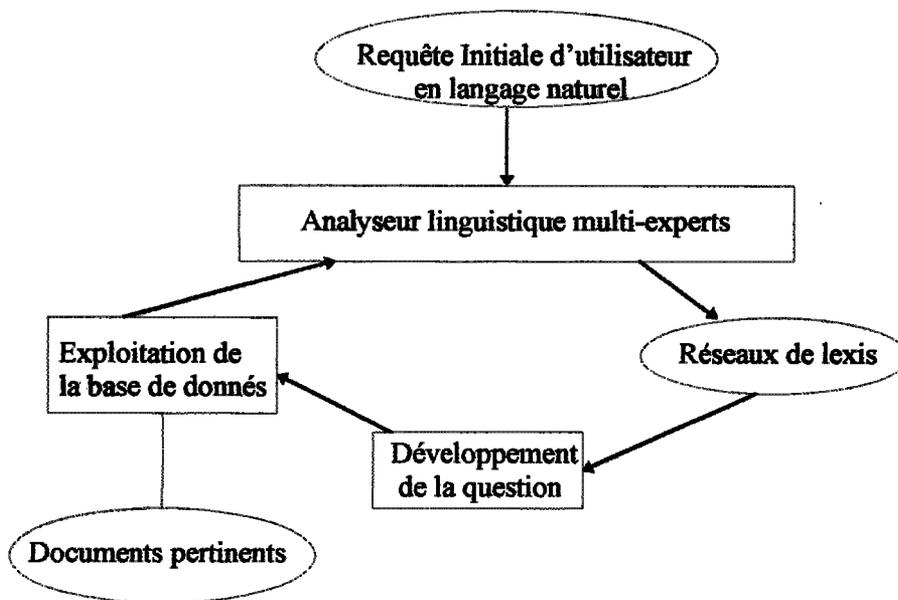


Figure 4 - Fonctionnement global du prototype de Mecabouche et Bassano

c) Prototype utilisant les architectures neuro-mimétiques et les systèmes multi-experts

Ce prototype part de l'architecture classique multi-expert [BASSANO] et propose de « casser » les différents experts définis dans l'architecture multi-experts en un ensemble de micro-experts. Tous ces micro-experts seront gérés par un réseau neuronal, utilisant l'apprentissage type, un ensemble d'entraînement et le traitement de pondérations avec algorithme d'entropie. Comme l'approche de MECABOUCHE et BASSANO, ce prototype considère la question de la reformulation des requêtes [DESROCQUES et al.].

2.4 Représentations du contenu et évaluation des réponses

La typologie des SRI par rapport à la façon de représentation du contenu des documents dans une bases de données est étroitement liée à la caractéristique de l'application.

Pour établir la typologie des SRI on va restreindre l'univers d'application à ceux qui traitent des données textuelles, soit composées par des références, soit de la propre information primaire. Selon les documents trouvés dans la littérature spécialisée, on s'est aperçu que les systèmes utilisent les approches suivantes : a) Représentation par mots ou groupe de mots ; b) Représentation utilisant des bases de connaissances ; c) Représentation par structure syntaxique.

2.4.1 Représentation par mots ou groupe de mots

Un Système de Recherche d'Information typique sélectionne les documents d'une base de données en réponse aux requêtes des utilisateurs et les présente selon un ordre de pertinence. La procédure habituelle consiste d'une part d'une méthode d'indexation afin de sélectionner les termes (mots ou phrases), et d'autre part en la création d'un fichier inversé d'indice lequel permet l'accès facile aux documents [STRZALKOWSKI] .

L'approche conventionnelle pour la représentation du contenu des documents est réalisée soit par l'indexation automatique du texte par mots ou racines de mots soit par l'indexation par mots ou phrases extraits manuellement d'un vocabulaire contrôlé. [SMEATON²]

L'indexation automatique peut se faire selon plusieurs méthodes. On peut distinguer :

- a) des méthodes élémentaires comme la simple extraction des mots où ces mots sont des descripteurs en excluant les mots vides [FLUHR] ;

- b) des méthodes statistiques que consistent à définir un modèle probabiliste des occurrences des mots dans un document considéré à l'intérieur d'une collection bien définie de manière à établir leur caractère pertinent pour participer à la description du contenu [BOUCHE] ;
- c) des méthodes linguistiques où l'identification des parties du discours pertinentes se fait sur des critères linguistiques. Une des approches linguistiques part du principe que le lexique, en tant que composant de la langue, ne contient que des éléments qui sont des propriétés, c'est-à-dire des prédicats. Le mot est donc un prédicat et il ne peut pas être considéré de façon isolé car il fait référence à un objet de la réalité extra-linguistique de l'auteur du document. Il ne peut donc pas être un descripteur [BOUCHE].

La finalité du descripteur exclut qu'on puisse l'envisager en faisant abstraction de la valeur référentielle de ses occurrences dans le corpus. Les mots de la langue, en tant qu'ils sont mots de la langue, ne signifient que des attributs, et non des substances, tant qu'ils ne sont pas mis en œuvre dans le discours. Le descripteur, quant à lui signifie une entité au sens de la philosophie d'Aristote. Le descripteur ne peut donc pas être considéré, à l'instar des mots de la langue comme un symbole sans référence [LE GUERN¹].

Ainsi, un descripteur doit être représenté par la plus petite unité du discours porteuse d'une valeur référentielle, dénommé syntagme nominal. C'est elle qu'il importe d'identifier dans le document [BOUCHE].

Deux phénomènes linguistiques limitent la compétence de l'utilisateur en choisir les termes de recherche : a) la polysémie, mots avec multiples sens ; et b) la synonymie, plusieurs

mots ou phrases pour désigner le même concept. La polysémie réduit la « précision » - mesure de performance d'une recherche d'information qu'indique la proportion de documents trouvés qui sont réellement pertinents - tandis que la synonymie réduit le « rappel » - mesure de performance d'une recherche d'information qu'indique la proportion de documents pertinents qui sont vraiment trouvés [MAULDIN].

Pour la détermination de la similarité entre les termes de la requête et ceux des documents on a trois modèles : 1) comparaison exacte ; 2) espace vectoriel ; 3) modèles de probabilités. Le modèle de comparaison exacte s'applique plus à la recherche bibliographique, tandis que les modèles espace vectoriel et de probabilités présentent des meilleurs résultats lorsque ils sont utilisés sur les bases de données textes intégral [PRITCHARD-SCHOCH].

Exemples :

a) WIN - Westlaw Is Natural [PRITCHARD-SCHOCH]

WIN est un Système de Recherche d'Information en Langage Naturel sur la base WESTLAW, développé par Howard Turtle et Jim Olson. WESTLAW est une base de données (en texte intégral) juridique de l'entreprise WEST Publishing Company. Ce système manipule des requêtes en langage naturel. Les techniques utilisées ont été le traitement automatique du langage naturel pour le traitement de la requête et les techniques statistiques pour la détermination de la similarité entre la requête et les documents trouvés ;

b) SPIRIT

SPIRIT est un Système de Recherche d'Information en Langage Naturel multilingue réalisé par la société Systemx. Ce système travaille avec des bases de

données textuelles. Pour la représentation du contenu des documents de la base de données, il utilise un traitement linguistique et statistique avant de la mise à jour du fichier inversé. La procédure de recherche d'information est faite aux moyens de trois différents outils : 1) le traitement linguistique ; 2) le traitement statistique ; 3) le traitement de la reformulation de la requête (au cas où la recherche n'a pas trouvé ou a trouvé une quantité insuffisante de documents, une procédure de reformulation permettant d'augmenter la représentation de chaque concept de la requête). Cette reformulation est faite à partir du traitement au niveau lexicale et des règles spécifique d'un thesaurus. Les réponses données par le Spirit sont triées par ordre décroissante de pertinence [RADWAN et al.].

2.4.2 Représentation utilisant des bases de connaissances

Le langage naturel s'est montré insuffisant pour représenter ou décrire les faits dans un ordinateur. Sa richesse, sa variété et son énorme ambiguïté ont diminué son utilisation en tant que moyen de description des faits dans l'ordinateur [ALBERICO et MICCO].

Selon Wiedenhold, des données sont des faits objectifs que peuvent être vérifiés. Un compteur que enregistre le nombre de personnes qu'entrent et sortent d'un bâtiment est une collection de données. Par contre, les connaissances sont abstraites, subjectives et ne sont pas facilement vérifiées. Les connaissances viennent de l'éducation et de l'expérience, elles ne viennent pas directement du monde réel. Elles sont plus floues et générales que les données [WIEDENHOLD].

La littérature inclue une grande variété de différentes méthodes de représentation de connaissances, depuis les systèmes formels basés sur des calculs propositionnels et des prédicats qui définissent les liens logiques restreints entre des entités, jusqu'aux grandes structures de réseau qui décrivent les entités par différents types d'attributs et qui inclue plusieurs différents liens parmi les entités [SALTON].

Exemples :

a) SCISOR

C'est un système de recherche d'information conceptuel. Ce type de système traite un ensemble d'information textuelle et l'analyse créant une base de connaissances de manière à fournir des réponses directes, en langage naturel, aux requêtes des usagers, elles aussi en langage naturel. La différence de ce type de système par rapport aux autres systèmes traditionnels de recherche d'information, est que la réponse est toujours une information primaire et non une référence. Ce système utilise des outils de traitement automatique du langage naturel (analyseurs morphologique, syntaxique, sémantique) pour l'analyse des documents et des requêtes. Pour la représentation du contenu de ces documents il utilise une base de connaissances [SMEATON²].

b) NAT-MULTILING

C'est une interface de recherche d'information multilingue commercialisée par Cybernetix. Cette interface peut fonctionner avec des bases de données relationnelles (SQL) et avec des bases de données textuelles ou bibliographiques (TEXTO). L'approche est fondée sur la compréhension des phrases et la traduction de ses signifiés en formule logique de première ordre. Ainsi elle combine les outils de traitement automatique du langage naturel avec la programmation logique.

L'interface accepte la requête en langue française, anglaise et espagnole. Les requêtes formulées dans une de ces langues sont traduites vers un langage logique [CHOCH DI YACOVO et al.].

c) Prototype [HESS]

C'est un système de recherche d'information documentaire basée sur des principes linguistiques et logiques. Ce système essaye de remplir la lacune qu'existe entre la syntaxe et la sémantique en utilisant des outils développés par le traitement automatique du langage naturel. Il utilise une forme logique pour représenter le contenu d'un ensemble d'informations afin de permettre d'éliminer des problèmes d'ambiguïté.

2.4.3 Représentation par structures syntaxiques

Ce type de représentation privilégie la phrase comme moyen de comparaison. Son objectif est d'obtenir à partir de l'analyse linguistique des documents les structures syntaxiques qui sont extraites et organisées sous forme d'arbre ajoutés aux informations syntaxiques des unités de la phrase. Cette structure est utilisée dans la procédure de détermination de la similarité entre les documents et les requêtes.

Le justificatif pour ce type d'approche est qu'elle peut résoudre les problèmes de polysémie et de synonymie [MAULDIN].

Exemples de cette approche :

a) ESPRIT SIMPR

Ce projet utilise, pour la représentation du contenu des documents et des requêtes, une structure que s'appelle TSA - *Tree Structured Analitics*, planifiée pour codifier

les différents types d'ambiguïtés. Cette structure consiste en un arbre binaire où sont stockés des informations comme: mots d'origine dans la phrase, ses racines, étiquettes de catégorie lexicale, morphologique et syntaxique [SMEATON et SHERIDAN].

b) FERRET - Flexible Expert Retrieval of Relevant English Text

La construction du système FERRET est fondé sur la supposition de la pertinence homéomorphiste. Cette supposition établit que les opérations d'analyse de la requête, de la comparaison des structures *Case Frames* et d'analyse des documents forment un homéomorphisme avec le jugement de pertinence des humains.

Ce projet consiste en un système conceptuel de recherche d'information et utilise, de manière analogue au projet SIMPR, l'analyse linguistique pour extraire les concepts et les représenter sous forme d'une structure dénommée *Case Frames*, qu'a été dessiné pour résoudre les problèmes spécifiques d'ambiguïtés [MAULDIN].

CONCLUSION

On a vu dans ce travail la description de quelques SRI dits en langage naturel. Ils ont été classés selon les 4 typologies établies à partir de la littérature trouvée. En faisant cette classification on peut se rendre compte des approches mises en oeuvre dans les dernières années pour le développement de ces systèmes. Or, il est difficile d'établir une définition précise de tels systèmes. On arrive plutôt à obtenir les caractéristiques les plus souhaitables pour eux.

Améliorer la convivialité d'utilisation de ces systèmes et augmenter la précision des résultats de la recherche d'information sont les traits principaux. Ainsi on peut envisager les caractéristiques souhaitables suivantes :

- ◊ interface de dialogue en langage naturel.

Cette interface doit posséder une interaction avec l'utilisateur de manière facile et naturelle afin d'éviter les contraintes d'un apprentissage d'un langage avec des règles rigides de grammaire. Pour cela, il faut donc utiliser les éléments que les humains utilisent pour se communiquer les uns avec les autres. Autrement dit, il faut utiliser un langage plus proche possible du langage parlé pour les humains. Ainsi l'interface doit être construite vers l'utilisateur final ;

- ◊ méthode de traitement du contenu des documents et des requêtes en considérant la façon dont les humains reconnaissent la pertinence ou non d'un document.

C'est-à-dire, il ne suffit pas de faire la simple extraction de mots, groupes de mots ou structures syntaxiques pour représenter les contenus des documents. Mais,

principalement identifier « les parties plus petites du discours (documents) que peuvent servir de base à une relation référentielle autonome, les syntagmes nominaux » [LE GUERN²].

Les humains sont capables de distinguer en un coup d'oeil lorsqu'ils regardent un document, s'il est ou non pertinent par rapport à son besoin d'information. Certes derrière ce processus il y a des connaissances qu'ont été accumulées pendant des années. Or, lorsque les hommes font le choix des documents pertinents parmi les autres, en regardant soit les titres, soit le propre contenu, ils ne font que comparer les éléments référentiels existant dans ces champs avec ceux qu'ils ont définis comme leur besoins d'information ;

◊ mécanisme de retro-alimentation.

Pour améliorer encore la précision des résultats de la recherche d'information, on considère qu'un mécanisme d'aide aux usagers sera intéressant, pour les raisons suivants: a) permet un raffinement de la requête; b) ce type d'interaction entre les usagers et les systèmes, permettra aux usagers connaître meilleur la base de données et le domaine de ses documents. Il y a déjà quelques prototypes qui ont développé ce type de mécanisme, appelé reformulation [voir MECABOUCHE].

En outre, à mon avis l'utilisation de plusieurs modes d'expression peut rendre les SRI plus conviviaux et agréables pour les usagers.

Par contre les problèmes d'ambiguïtés, d'anaphore et d'ellipses continueront à exister, étant donné qu'ils sont inhérents à la langue naturelle. Ces problèmes ne peuvent pas être

résolus complètement. On envisage ici l'importance de la troisième caractéristique souhaitable (sur les mécanisme de retro-alimentation). C'est donc au moyen de cette interaction que les usagers peuvent faire un raffinement, en utilisant de son stock de savoir, et orienter les SRI vers un résultat plus précis et cohérent avec son besoin d'information.

Dans ce aspect là, on considère la troisième génération d'interface de dialogue, classification établie par [BINOT et al.²], comme une tendance, comme approche, pour le développement des Systèmes de Recherche d'Information en Langage Naturel.

REFERENCES BIBLIOGRAPHIQUES

[ADDISON et NELSON]

ADDISON, E. R. et NELSON, P. E. « Intelligent Hypertext ». *Proceedings of the Thirteenth National Online Meeting*. New York, 1993. p. 27-30.

[ALBERICO et MICCO]

ALBERICO, Ralph et MICCO Mary. *Expert Systems for Reference and Information Retrieval*. Westport : Meckler, 1990. 395 p. (Supplements to computers in libraries).

[BALPE]

BALPE, Jean-Pierre. *Hyperdocuments, Hypertextes, Hypermedia*. Paris : Eyrolles, 1990. 200 p.

[BASSANO]

BASSANO, Jean-Claude. *DIALECT: un système expert pour la recherche documentaire*, Paris, 1986, Thèse de doctorat d'Etat, Université Paris 11. ^{NC}

[BINOT et al.¹]

BINOT, J.-L., FALZON, P., PEREZ, R., PEROCHE, B., SHEEHY, N., ROUAULT J. et WILSON, M. « Architecture of a multimodal dialogue interface for knowledge-based systems ». *Actes de la Conférence ESPRIT 90*. Novembre, 1990. ^{NC}

[BINOT et al.²]

BINOT J.-L., DEBILLE, L., SEDLOCK, D., VANDECAPELLE, B. « Représentation Sémantique et Interprétation dans une Interface en Langage Naturel ». *Le Français Moderne*. Juin, 1991, t. LIX, n°. 1. p. 57-84.

[BOUCHE]

BOUCHE, Richard. « Le Syntagme Nominal, une Nouvelle Approche des Bases de Données Textuelles ». *Meta*. 1989, vol. 34, n°. 3. p. 428-434.

[CLITHEROW et al.]

CLITHEROW, Peter, RIECKEN, Doug et MULLER, Michael. « VISAR : A System for Inference and Navigation in Hypertext ». *Hypertext'89. Proceedings. Special Issue - SIGCHI BULLETIN*. Pittsburgh, November 5-8, 1989. p. 293-304. ^{NC}

[COCH DI YACOVO et al.]

COCH DI YACOVO, José, CRISPINO, Gustavo, CUKIERMAN, Diana, MORIZE, Geneviève et WONSEVER, Dina. « NAT-MULTILING, tools for multilingual interfaces with data bases ». *RIA0 91 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991, vol. 1. p. 514-525.

[CORET et al.]

CORET, Annie, MENON, Bruno, SCHIBLER, Daniele et TERRASSE, Christophe. « Un système d'indexation structurée à l'INIST. Bilan d'une étude préalable ». *Documentaliste - Sciences de l'Information*. 1994, vol. 31, n°. 3. p. 148-158.

^{NC} Non Consultée

[DAMI et LALLICH-BOIDIN]

DAMI, Samir et LALLICH-BOIDIN, Geneviève. « An Expert System for French Analysis within a Multi-mode Dialogue to be Connected ». *RIAO 91 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991. vol. 1. p. 431-451.

[DEWEZE]

DEWEZE, A. *Informatique documentaire*. 4^e éd. Paris : Masson, 1993. 292 p.

[DESROCQUES et al.]

DESROCQUES, Gilles, BASSANO, Jean-Claude et ARCHAMBAULT, Dominique. « An Associative Neural Expert System for Information Retrieval ». *RIAO 91 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991. vol. 1. p. 546-566.

[FLUHR]

FLUHR, Christian. « Le traitement du langage naturel dans la recherche d'information documentaire ». In.: *Cours INRIA - Interfaces Intelligentes dans l'Information Scientifique et Technique*. 18-22 Mai 1992. p. 103-128.

[GUIMIER-SORBETS]

GUIMIER-SORBETS, Anne-Marie. « Des textes aux images. Accès aux informations multimédias par le langage naturel ». *Documentaliste - Sciences de l'Information*. 1993, vol. 30, n^o. 3. p. 127- 134.

[HARTER]

HARTER, Stephen P. *Online Information Retrieval : Concepts, Principles and Techniques*. Orlando : Academic Press. Inc., 1986. 259 p. (Library and Information Science).

[HAUSSER]

HAUSSER, R. *Computation of Language*. Springer, Berlin, 1989. ^{NC}

[HESS]

HESS, Michael. « An Incrementally Extensible Document Retrieval System Based on Linguistic and Logical Principles ». *SIGIR'92*. Denmark, June, 1992. p. 190-197.

[KAHLE et al.]

KAHLE, Brewster, MORRIS, Harry, DAVIS, Franklin, ERICKSON, Thomas, HART, Clare et PALMER, Robin. « Wide Area Information Servers : An Executive Information System for Unstructured Files ». *International High Speed Networks for Scientific and Technical Information (AGARD)*. 1993. p. 12/1-9.

[KONG et SHOW]

KONG, Hinny et SHOW, Guan Yeong. « Evaluation of parsing techniques for natural language processing ». *Proceedings of the International Conference on Information for natural language processing*. Singapore, 1991. p. 422-432.

[KUPIEC]

KUPIEC, J. « MURAX: a robust linguistic approach for question answering using an on-line encyclopedia ». *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Pittsburgh, 1993. p. 181-190.

^{NC} Non Consultée

[LANCEL]

LANCEL, Jean Marie. « A pragmatic-based language understanding system (PLUS) ». *CEC-ESPRIT (Information Processing System)*. November, 1990.

[LE GUERN¹]

LE GUERN, Michel. « Les descripteurs d'un système documentaire : essai de définition », In : Bès, G.C., Fauchère, P.M., Lagueunière, F. *Actes du Colloque « Traitement automatique des langues naturelles et systèmes documentaires »*. Condenser, supplément I, Université Clermont Ferrand, 1982. ^{NC}

[LE GUERN²]

LE GUERN, Michel. « Un analyseur morpho-syntaxique pour l'indexation automatique », *Le Français Moderne*. Juin, 1991, t. LIX, n°. 1, p. 22-35.

[MAREGA et PAZIENZA]

MAREGA, Roberto et PAZIENZA, Maria Teresa. « CoDHIR : an information retrieval system based on semantic document representation ». *Journal of Information Science*. 1994, vol. 20, n°. 6. p. 399-412.

[MAULDIN]

MAULDIN, Michael L. « Retrieval Performance in FERRET : A Conceptual Information Retrieval System ». *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*. Chicago, October 13-14, 1991. p. 347-355.

[MEKABOUCHE et BASSANO]

MEKABOUCHE, A. et BASSANO, Jean-Claude. « Multi-experts Systems for Documentary research ». *RIAO 91 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991. vol. 1, p. 394-413.

[NIE]

NIE, Jian-Yun. « Towards a Probabilistic Modal Logic for semantic-based Information Retrieval ». *15th Annual International SIGIR*. 1992. p. 140-151.

[POLITY]

POLITY, Yolla. « Evaluation des modes de recherche en langage naturel ». *Documentaliste - Sciences de l'Information*. 1994, vol. 31, n°. 3. p. 136-142.

[POLLITT]

POLLITT, Steven. « CANSEARCH : An expert systems approach to document retrieval ». *Information Processing and Management*. 1987, vol. 23, n°. 2. p. 119-138.

[PRITCHARD-SCHOCH]

PRITCHARD-SCHOCH, Teresa. « Natural Language Comes of Age ». *Online*. 1993, May, p. 33-43.

^{NC} Non Consultée

[RADWAN et al.]

RADWAN, Khaled, FOUSSIER, Frédéric et FLUHR, Christian. « Multilingual Access to Textual Databases ». *RIAO 91 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991. vol. 1. p. 475-489.

[ROWE et GUGLIELMO]

ROWE, Neil C. et GUGLIELMO, Eugene J. « Exploiting Captions in Retrieval of Multimedia Data ». *Information Processing & Management*. 1993, vol. 29, n°. 4. p. 453-461.

[ROWLEY]

ROWLEY, Jennifer. « The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research ». *Journal of Information Science*. 1994, vol. 20, n°. 2. p. 108-119.

[SALTON et MCGILL]

SALTON, Gerard et MCGILL, Michael J. *Introduction to modern information retrieval*. New York : McGraw-Hill Book Company, 1983. 448 p. (Computer Science).

[SALTON]

SALTON, Gerard. *Automatic Text Processing : The Transformation, Analysis and Retrieval of Information by Computer*. Massachusetts : Addison-Wesley Publishing Co., 1989. 530 p. (Computer Science).

[SAMSTAG-SCHNOCK et MEADOW]

SAMSTAG-SCHNOCK, Uwe, MEADOW, Charles T. « PBS : An Economical Natural Language Query Interpreter ». *Journal of the American Society for Information Science*. 1993, vol. 44, n°. 5. p. 265-272.

[SHOW et al.]

SHOW Guan Yeong, KONG, Hanny et LIN, Kenneth Wente. « Intelligent user interface to SQL-based database system ». *Engineering Application Artificial Intelligence*. 1993, vol. 6, n°. 4. p. 307-316.

[SMEATON¹]

SMEATON Alan F. « Information retrieval and natural language processing ». *Informatics 10: prospects for intelligent retrieval: proceedings of a conference jointly sponsored by ASLIB*. University of York, 21-23 Mars, 1989. p. 1-14.

[SMEATON²]

SMEATON Alan F. « Prospects for intelligent, language-based information retrieval ». *Online Review*. 1991, vol. 15, n°. 6. p. 373-382.

[SMEATON et SHERIDAN]

SMEATON, Alan F. et SHERIDAN, Paraic. « Using Morpho-Syntaxique Language Analysis in Phrase Matching ». *RIAO 9 : Recherche d'Information Assistée par Ordinateur*. Barcelona, 1991. vol. 1, p. 414-430.

[STRZALKOWSKI]

STRZALKOWSKI, Tomek. « Natural language processing in large-scale text retrieval tasks ». *Text REtrieval Conference (TREC-1)*. Gaithersburg, 1993. p. 173-187.

[WANG]

WANG, Fangju. « Towards a natural language user interface : an approach of fuzzy query ». *International Journal of Geographical Information Systems*. 1994, vol. 8, n°. 2. p. 143-162.

[WIEDENHOLD]

WIEDENHOLD, Gio. « Structural versus application knowledge for improved database interfaces ». In. : *Second Conference on Computer Interfaces and Intermediaries for Information Retrieval*. Alexandria, Defense Technical Info. Ctr., 1986. p.17-96.^{NC}

BIBLIOGRAPHIE

[ADDISON et al.]

ADDISON, E. R., WILSON, H. D. et FEDER, J. « The impact of plain English searching on end users ». *Proceedings of the Fourteenth National Online Meeting*. New York, 1993. p. 5-9.

[AMARAL et SATOMURA]

AMARAL, M. B. do et SATOMURA, Y. « Processing natural languages at Chiba University Hospital ». *M Computing*. 1993, vol. 1, n° 4. p. 6, 9-15.

[ARCHAMBAULT et BASSANO]

ARCHAMBAULT, D. et BASSANO, Jean-Claude. « A neural network for supervised learning of natural language grammar ». *Proceedings of the Sixth International Conference on Tools with Artificial Intelligence. TAI 94*. New Orleans, 1994. p. 267-273.

[BASCH]

BASCH, R. « Searching in plain English ». *Link-Up*. USA, 1994, vol. 11, n° 2. p. 14-15.

[BELKIN et al.]

BELKIN, (N. J.), COOL, C., CROFT, W. B., CALLAN, J. P. « The effect of multiple query representations on information retrieval system performance ». *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Pittsburgh, 1993. p. 339-347.

[BJORNER]

BJORNER, S. « The . Where And . Why of FREESTYLE ». *Online*. 1994, vol. 18, n° 2. p. 88-91.

[BLAIR]

BLAIR, D. C. « Information retrieval and the philosophy of language : information retrieval ». *Computer Journal*. 1992, vol. 35, n° 3. p. 200-207.

[BRAUNWARTH et al.¹]

BRAUNWARTH, M., MEKABOUCHE, A. et BASSANO, Jean-Claude. « DIALECT-2 : an information retrieval system based on distributed artificial intelligence tools ». *Proceedings of the Sixth International Conference on Tools with Artificial Intelligence. TAI 94*. New Orleans, 1994. p. 800-803.

[BRAUNWARTH et al.²]

BRAUNWARTH, M., MEKABOUCHE, A. et BASSANO, Jean-Claude. « Using a dynamic blackboard model in a documentation retrieval system ». *Proceedings of Avignon '93. 13th International Conference*. Avignon, 1993. vol. 3. p. 33-44.

[CARACENI et al.]

CARACENI, R., GRAZIADIO, B., MUSSETTO, P., OBLIEGHT, A. et ZINNO, A. « Integrating data and text retrieval in a natural language system ». *Proceedings of Avignon '93. 13th International Conference*. Avignon, 1993. vol. 3. p. 55-64.

[CARENINI et al.]

CARENINI, G., PIANESI, F., PONZI, M. et STOCK, O. « Natural language generation and hypertext access ». *Applied Artificial Intelligence*. 1993, vol. 7, n° 2. p. 135-164.

[CROFT]

CROFT, W. B. « The university of Massachusetts TIPSTER project ». *NIST special publication - TREC-1: Text Retrieval Conference*. Gaithersburg, 1992. p. 101-105.

[DE BERTRAND DE BEUVRON]

DE BERTRAND DE BEUVRON, François. Un système de progamation logique pour la création d'interfaces homme-machine en langue naturelle. 1992. 293 p. Thèse de doctorat d'Etat, Université de Compiègne.

[DE BRITO¹]

DE BRITO, Marcilio. Réalisation d'un analyseur morpho-syntaxique pour la reconnaissance du syntagme nominal : utilisation des grammaires affixes. Lyon, 1991. 221 p. Thèse de doctorat d'Etat, Université Claude Bernard, Lyon I.

[DE BRITO²]

DE BRITO, Marcilio. « Information System in natural languages: the search for an automatic indexing system ». *Ciência da Informação*. 1992, vol. 21, n° 3. p. 223-232.

[DECLES]

DECLES J. P. « Les interfaces homme-machine en langage naturel dans l'interrogation des bases de données relationnelles ». *IDT. Information, documentation, transfert des connaissances*. Paris, 1994. p. 89-93.

[DESERT]

DESERT, S. E. « WESTLAW is Natural v. Boolean searching: a performance study ». *Law Library Journal*. 1993, vol. 85, n° 4. p. 713-742.

[DRISCOLL et al.]

DRISCOLL, J., LAUTENSCHLAGER, J. et MIMI, Zhao. « The QA system », *NIST special publication - TREC-1: Text Retrieval Conference*. Gaithersburg, 1992. p. 199-207.

[EVANS]

EVANS R. « Beyond boolean: relevance ranking, natural language and the new search paradigm ». *Proceedings of the Fifteenth National Online Meeting*. New York, 1994. p. 121-128.

[EVANS et LEFFERTS]

EVANS, D.A. et LEFFERTS, R. G. « Design and evaluation of the CLARIT-TREC-2 system ». *TREC-1: Text Retrieval Conference*. Gaithersburg, 1992. p.137-150.

[GALICY et al.]

GALICY Jean Pierre, JOUIS, Christophe et GRAU, Brigitte. SILNEBAD : Système d'interrogation en langage naturel d'un ensemble de bases de données. 1989. 85 p. Rapport.

[GIROLLET et VICTORRI]

GIROLLET, D. et VICTORRI, B. « The linguistic analyser of a smart gateway ». *SEPLN - Sociedad Espanola para el Processamiento del Lenguaje Natural. VIII Congreso*. Granada, 1993. p. 29-39.

[GUARDALBEN et LUCARELLA]

GUARDALBEN, G. et LUCARELLA, D. « Information retrieval based on fuzzy reasoning »; *Data & Knowledge Engineering*. 1993, vol. 10, n° 1. p. 29-44.

[GRIFFITH]

GRIFFITH, C. « Westlaw's WIN: not only natural, but new ». *Information Today*. 1992, vol. 9, n° 9. p. 9-11.

[HERSH]

HERSH, W. R. « SAPHIRE: a concept-based approach to automated indexing and retrieval in the biomedical domain ». *CURRENT RESEARCH 1992 National Library of Medicine Grant*.

[JACSO]

JACSO, P. « Don't kiss Boolean goodbye. It's AND not OR, let alone XOR ». *Information Today*. 1994, vol. 11, n° 2. p. 22-24.

[LAROUK]

LAROUK, Omar. « Linguistico-statistical approach and logic applied in documentary system ». *Proceedings of 8th SIGAPP Symposium on Applied Computing*. Indianapolis, 1993. p. 737-744.

[LARSON]

LARSON, Ray R. « Classification clustering, probabilistic information retrieval, and the online catalog ». *Library Quarterly*. 1991, vol. 61, n° 2. p. 133-173.

[LASSALLE]

LASSALLE, E. « Telmi: a reusable information retrieval system and its applications ». *ASLIB Proceedings*. 1993, vol. 45, n° 5. p. 144-148.

[LE LOARER]

LE LOARER, Pierre. « OPAC: opaque or open, public, accessible and co-operative; some developments in natural language processing ». *Program*. 1993, vol. 27, n° 3. p. 251-268.

[LEIGH]

LEIGH, Sharon A. « The use of natural language processing in the development of topic specific databases ». *Proceedings of the Twelveth National Online Meeting*. New York, 1991. p. 209-213.

[LIDDY et al.]

LIDDY, E.D., WOOJIN, Paik et YU, E. S. « Text categorization for multiple users based on semantic features from a machine-readable dictionary ». *ACM Transactions on Information System*. 1994, vol. 12, n° 3. p. 278-295.

[LIDDY]

LIDDY, E. D. « An alternative representation for documents and queries ». *Proceedings of the Fourteenth National Online Meeting*. New York, 1993. p. 279-284.

[LOSEE]

LOSEE, R. M., Jr. « Term dependence: truncating the Bahadur Lazarsfeld expansion ». *Information Processing & Management*. 1994, vol. 30, n° 2. p. 293-303.

[MCCRAY et al.]

MCCRAY, A. T., ARONSON, A. R., BROWNE, A. C., RINDFLESCH, T. C. et SRINIVASIN, A. « UMLS knowledge for biomedical language processing ». *Bulletin of the Medical Library Association*. 1993, vol. 81, n° 2. p. 184-194.

[LUBKOV]

LUBKOV, M. « Lexic: le langage naturellement ». *Archimag*. 1994, n° 74. p. 33-35.

[MEKABOUCHE]

MEKABOUCHE, A. Un système multi-experts pour la recherche documentaire. France, 1991. 263 p. Thèse de doctorat d'Etat, Université d'Orleans.

[MENDIBOURE]

MENDIBOURE, Catherine. DOHQL: Une interface avancée pour l'interrogation de bases de données orientées-objet et documentaires. Toulouse, 1994. 227 p. Thèse de doctorat d'Etat, Université de Toulouse 3.

[METZGER]

METZGER, J-P. Syntagmes Nominaux et information textuelle : reconnaissance automatique et représentation. Lyon, 1988. 324 p. Thèse de doctorat d'Etat, Université Claude Bernard, Lyon I.

[NELSON¹]

NELSON, P.E. « The ConQuest system ». *Text REtrieval Conference (TREC)*. Gaithersburg, 1993. p. 265-270.

[NELSON²]

NELSON, P.E. « Site report for the Retrieval Conference ». *NIST special publication - TREC-1 : Text REtrieval Conference (TREC)*. Gaithersburg, 1992. p. 287-296.

[OTT]

OTT, N. « Aspects of the automatic generation of SQL statements in a natural language query interface ». *Information Systems*. 1992, vol. 17, n° 2. p. 147-159.

[OWEI et HIGA]

OWEI, V. et HIGA, K. « A paradigm for natural language explanation of database queries: a semantic data model approach ». *Journal of Database Management*. 1994, Winter, vol. 5, n° 1. p. 18-30.

[PERNEL]

PERNEL, Didier. Gestion des buts multiples de l'utilisateur dans un dialogue homme-machine de recherche d'informations. Paris, 1994. 328 p. Thèse de doctorat d'Etat, Université de Paris 11.

[PRITCHARD-SCHOCH]

PRITCHARD-SCHOCH, Teresa. « WIN - Westlaw goes natural ». *Online*. 1993, Jan, vol. 17, n° 1. p. 101-103.

[PUGET]

PUGET, Dominique. Aspects sémantiques dans les systèmes de recherche d'informations. Toulouse, 1993. 183 p. Thèse de doctorat d'Etat, Université de Toulouse 3.

[PURDY]

PURDY, W. C. « Alogic for natural language ». *Notre Dame Journal of Formal Logic*. 1991, vol. 32, n° 3. p. 409-425.

[RABEN]

RABEN, J. « SCHOLAR: implication for business ». *Proceedings of the Fourteen National Online Meeting*. New York, 1993. p. 343-348.

[RAMMAL]

RAMMAL, Mahmoud. Une interface conceptuelle pour le traitement du langage naturel. Application au langage médical dans le système ADM. 1993. 202 p. Thèse de doctorat d'Etat, Université de Compiègne.

[ROWE]

ROWE, Neil C. « Integrating depictions in natural-language captions for efficient access to picture data ». *Information Processing & Management*. 1994, vol. 30, n° 3. p. 379-388.

[SANDOVAL et al.]

SANDOVAL, A. M., MORENO, C. O., GRISHMAN, R., MACLEOD, C. et STERLING, J. « PROTEUS: a multilingual system for information retrieval ». *SEPLN - Sociedad Espanola para el Processamiento del Lenguaje Natural. VIII Congreso*. Granada, 1993. p. 47-56.

[SHULDBERG et al.]

SHULDBERG, H. K., MACPHERSON, M., HUMPHREY, P. et CORLEY, J. « Distilling information from the EDS Template Filler System ». *Journal of the American Society for Information Science*, 1993. vol. 44, n° 9. p. 493-507.

[SMEATON³]

SMEATON, Alan F. « Progress in the application of natural language processing to information retrieval tasks : information retrieval ». *Computer Journal*. 1992, vol. 35, n° 3. p. 268-278.

[SUTCLIFFE]

SUTCLIFFE, R. F. E. « PELICAN : a prototype information retrieval system using distributed propositional representation ». *Proceedings of AI and Cognitive Science '91*. Cork, 1991. p. 147-163.

[TECNOPIR]

TECNOPIR, C. « Online databases : natural language searching with WIN ». *Library Journal*. 1993, vol. 118, n° 18. p. 54-56.

[THOMAZEAU]

THOMAZEAU, Jacques. Une interface multimodale pour l'interrogation d'une base d'objets complexes et documentaires. Toulouse, 1993. 218 p. Thèse de doctorat d'Etat, Université de Toulouse 3.

[TRABELSI et al.]

TRABELSI, Z., KOTANI, Y., TAKIGUCHI, N. et NISIMURA, H. « A database-domain hierarchy-based technique for handling unknown terms in natural language database query

interfaces ». *IEICE Transactions on Information and Systems*. 1993, vol. E76-D, n° 6. p. 668-679.

[TURNER et al.]

TURNER, W. A., BUFFET, P. et LAVILLE, F. « LEXITRAN for an easier public access to patent database ». *World Patent Information*. 1991, vol. 13, n° 2. p. 81-90.

[VEGA et OGONOWSKI]

VEGA, J. et OGONOWSKI, A. « Natural-language access to the Dianeguide database », *SEPLN - Sociedad Espanola para el Processamiento del Lenguaje Natural. VIII Congreso*. Granada, 1993. p. 63-73.

[VICKERY et VICKERY]

VICKERY, B. et VICKERY, A. « An application of language processing for a search interface ». *Journal of Documentation*. 1992, vol. 48, n° 3. p. 255-275.

[VIEIRA]

VIEIRA, Simone Bastos. *La recuperación automática de información jurídica: metodología de análisis lógico-sintáctico para la lengua portuguesa*. Madrid, 1994. 383 p. Thèse de doctorat, Universidad Complutense de Madrid.

[YOUNG et al.]

YOUNG, Charlene W., EASTMAN, Caroline, M. et OAKMAN, Robert L. « An analysis of ill-formed input in natural language queries to document retrieval systems ». *Informaion Processing & Management*. 1991, vol. 27, n° 6. p. 615-622.

[ZAMPOLLI et CALZOLARI]

ZAMPOLLI, A. et CALZOLARI, N. « Linguistic tools for information retrieval. Documentary languages and databases ». *Advances in knowledge organization*. Rome, 1991. p. 174-200.

[ZWEIGENBAUM]

ZWEIGENBAUM, P. « MENELAS: an access system for medical records using natural language ». *Computer Methods and Programs in Biomedicine*. 1994, vol. 45, n° 1-2. p. 117-120.

