

**D.E.A. SCIENCES DE L'INFORMATION ET DE LA
COMMUNICATION**

**CARACTERISATION LINGUISTIQUE DES QUESTIONS
D'UTILISATEURS D'OPACs**

**Recherche à partir d'un corpus de sessions d'utilisateurs
recueillies sur l'opac de la Médiathèque de La Villette**

Nathalie RAMEL

**sous la direction de
Sylvie LAINE-CRUZEL et Mohamed HASSOUN**

**ECOLE NATIONALE SUPERIEURE DES SCIENCES DE L'INFORMATION ET
DES BIBLIOTHEQUES**

OPTION 3

1993

**D.E.A. SCIENCES DE L'INFORMATION ET DE LA
COMMUNICATION**

UNIVERSITE LYON 2

UNIVERSITE LYON 3

ENSSIB

**CARACTERISATION LINGUISTIQUE DES QUESTIONS
D'UTILISATEURS D'OPACs**

**Recherche à partir d'un corpus de sessions d'utilisateurs
recueillies sur l'OPAC de la Médiathèque de La Villette**

Nathalie RAMEL

**Sous la direction de
Sylvie LAINE-CRUZEL et Mohamed HASSOUN**

1993

SOMMAIRE

Volume 1

Introduction	1
Chapitre 1 - Traitement de la requête en langue naturelle	3
I- Le traitement linguistique de la requête	4
1- Typologie des catalogues en lignes	5
2- Modèles de recherche documentaire	6
2-1 Différents modèles de recherche documentaire	6
2-2 Le modèle booléen et ses limites	6
2-2-1 Le modèle booléen standard	6
2-2-2 Limites du modèle booléen et des systèmes de recherche "classiques"	7
2-3 Un OPAC de la seconde génération : l'OPAC de la Médiathèque de La Villette	8
3- Les systèmes s'attachant au traitement de la question	9
3-1 Les interfaces	9
3-1-1 Quelques réalisations	9
3-1-2 Critique de ces systèmes	10
3-2 Les systèmes traitant de la même manière documents et questions	11
3-3 Les systèmes experts de recherche documentaire	12
II- Le traitement linguistique au service des utilisateurs	14
1- Linguistique et convivialité des systèmes	14
1-1 Etudes sur les langages documentaires : distance et niveau linguistiques	14
1-2 Techniques de navigation dans les OPACs	15
1-3 La reformulation	16
1-3-1 Définition et méthodes	16
1-3-2 Quelques systèmes proposant une reformulation	17
2- Langage de l'utilisateur et modèles mentaux	18
III- Une recherche exploratoire : le programme PARINFO-OPACs	19
1- Le programme PARINFO	19
2- Le projet PARINFO-OPACs	19
2-1 Objectifs	19
2-2 Les participants au projet	20
2-3 Terrain d'enquête	21
2-4 Méthodologie	21
2-4-1 Trois phases	21
2-4-2 Mise en oeuvre de l'enquête	21
2-4-3 Perspectives	22

3-	Composante linguistique du projet.....	23
3-1	Objectifs.....	23
3-2	Première analyse des données.....	23
3-3	Prolongements à l'analyse linguistique.....	24
IV-	Modélisation d'une question en langage naturel.....	25
1-	Typologie linguistique de la question P011.....	25
1-1	Types de réponses.....	26
1-2	Types de recherches.....	27
1-3	Typologie sémantique.....	28
1-4	Analyse morpho-lexicale.....	29
1-4-1	Nombre de mots par réponse.....	29
1-4-2	Catégories morpho-lexicales.....	29
1-5	Analyse syntaxique.....	31
1-5-1	Catégories syntaxiques.....	31
1-5-2	Quelques phénomènes syntaxiques.....	32
2-	Comparaison PR24 / P011.....	34
2-1	Identité ou inclusion.....	35
2-2	Reprise de termes.....	36
2-3	Lien sémantique.....	37
Chapitre 2 -	Formulation de la requête, du préquestionnaire	
	au postquestionnaire.....	39
I-	Méthodologie.....	40
1-	Constitution d'un corpus.....	40
1-1	Critère de sélection.....	40
1-2	Sessions non retenues.....	40
1-3	Organisation des données.....	41
2-	Obstacles et limites.....	42
3-	Corpus et méthode d'étude.....	43
3-1	Quelques définitions.....	43
3-2	Typologie des recherches et des utilisateurs.....	44
3-3	Méthode d'étude.....	45
II-	Evolution de la formulation,	
	du préquestionnaire à la première requête.....	46
1-	Typologie des premières requêtes.....	46
2-	Comparaison PR24 / première requête par titre.....	47
3-	Comparaison PR24 / première requête par auteur.....	48
4-	Comparaison PR24 / première requête par auteur-titre.....	48
5-	Comparaison PR24 / première requête par mot(s) du titre, de l'auteur ou du sujet.....	49
6-	Comparaison PR24 / première requête par sujet.....	50
6-1	Typologie syntaxique.....	50
6-2	Typologie sémantique.....	51
6-3	De PR24 à la première requête par sujet.....	52
6-3-1	Identité.....	52
6-3-2	Inclusion du sujet dans la réponse PR24.....	53
6-3-3	Abandon de la formulation initiale.....	54
6-3-4	Un terme ou un syntagme commun.....	55
	Synthèse.....	55

III- Etude de la formulation durant le déroulement de la session.....	57
1- Problèmes d'utilisation des commandes.....	57
2- Recherche en une étape et un mode de recherche.....	59
2-1 Caractérisation des recherches.....	59
2-2 Modèles de recherche.....	59
2-3 Stratégies de recherche.....	61
3- Recherches en plusieurs étapes et un mode de recherche.....	62
3-1 Caractérisation des recherches.....	62
3-2 Modèles de recherche.....	63
3-3 Evolution sémantique.....	65
3-4 Instanciación.....	67
4- Recherches en plusieurs étapes et plusieurs modes de recherche.....	68
4-1 Caractérisation des recherches.....	68
4-2 Mode de recherche et évolution de la formulation.....	70
4-3 Instanciación.....	71
5- Sessions composées de recherches indépendantes.....	72
5-1 Recherches en une étape et un mode de recherche.....	73
5-2 Recherches en plusieurs étapes et un mode de recherche.....	73
5-3 Recherches en plusieurs étapes et plusieurs modes de recherche.....	74
IV- Synthèse.....	76
1- Tableau récapitulatif.....	76
2- Quelques stratégies de recherche.....	76
3- Question P011 et reformulation.....	77
3-1 Identité ou inclusion.....	78
3-2 Reprise de termes.....	78
3-3 Absence de lien.....	79
Conclusion.....	80
Chapitre 3 - Linguistique et protocoles expérimentaux.....	82
I- Apports et limites du questionnaire électronique.....	83
1- Intérêt des questionnaires électroniques.....	83
2- Linguistique et questionnaires.....	84
2-1 Aspects linguistiques dans les deux questionnaires d'enquête.....	84
2-2 Exploitation de la question P011.....	84
2-3 Interprétation des sessions.....	85
3- Limites du postquestionnaire.....	86
II- Quelques réorientations.....	88
1- Pour un postquestionnaire oral.....	88
2- Quelle population ?.....	88
2-1 Le public de la Médiathèque de La Villette.....	88
2-2 Typologie des usagers.....	89
2-3 Panel d'utilisateurs.....	90

III- L'évaluation par résolution de problèmes	92
1- Analyse de quelques dispositifs.....	92
1-1 Représentation mentale d'un système de recherche d'information.....	92
1-2 Recherche par reformulation.....	93
1-3 Enquête expérimentale sur deux types d'instruction pour les usagers de catalogues informatisés.....	93
2- Vers un nouveau dispositif ?.....	94
Conclusion	96
Bibliographie	98

J'adresse tous mes remerciements à

Mmes Sylvie Lainé-Cruzel, Maître de Conférences, ENSSIB
Danielle Roger, Conservateur de bibliothèque, ENSSIB
Maria Witt, Conservateur, Médiathèque de la Cité des
Sciences et de l'Industrie

et à MM. Mohamed Hassoun, Maître de Conférences, ENSSIB
Pierre Le Loarer, Directeur de projet, GSI-ERLI

qui m'ont communiqué les données de l'enquête et fait partager leur
expérience du projet PARINFO-OPACs.

Ce D.E.A. fait suite à un Diplôme de Conservateur de Bibliothèque
préparé à l'ENSSIB. Je souhaiterais donc remercier toutes les personnes et
instances qui ont soutenu mon projet de poursuivre ma formation dans cet
établissement, et plus particulièrement

Mme Marie-Jeanne Tinland, Directrice du Service Commun de la Documentation
de l'Université Lumière Lyon 2, dans lequel je prendrai mes fonctions de
Conservateur en Octobre 93,

M. Richard Bouché, Professeur à l'ENSSIB

et M. Jacques Kériguy, Directeur de l'ENSSIB.

INTRODUCTION

Toutes les études s'accordent sur ce point : les catalogues en ligne, les OPACs, ne satisfont pas leurs usagers et ne remplissent pas les missions pour lesquelles ils ont été conçus, à savoir permettre l'accès de tous à l'information bibliographique, et grâce à cette information, au fonds documentaire d'un établissement donné.

Alors que la plupart des bibliothèques préparent leur informatisation, et que l'on parle de mise en réseau des collections à travers des projets tels que le Catalogue Collectif de France, ce constat n'est plus acceptable.

Depuis une dizaine d'années, une nouvelle perspective se dessine, qui consiste à ne plus se contenter de progrès techniques, mais à intégrer une nouvelle dimension : celle de l'utilisateur, avec ses difficultés, ses blocages, mais aussi son intérêt pour les OPACs. Elle élabore études et enquêtes destinées à prendre la mesure des problèmes et à modéliser cette variable jusqu'alors laissée pour compte : le comportement de l'utilisateur face à la machine.

La recherche exploratoire sur les catalogues en ligne accessibles par le public lancée en 1991 dans le cadre du projet PARINFO (Programme d'Aide à la Recherche en Information) s'inscrit dans cette voie. Elle constitue même un tournant pour la recherche française sur les OPACs, puisqu'elle aboutit à la constitution d'un corpus de données analysables dans un contexte d'utilisation "réelle", la Médiathèque de la Cité des Sciences et de l'Industrie de La Villette. On ne disposait pas, jusqu'alors, d'un tel matériau.

L'enquête s'est aussi donné pour objectif de tracer de nouvelles pistes de recherche dans le domaine des OPACs, des projets pluridisciplinaires, qui font appel aux sciences de l'information, à l'informatique, aux sciences cognitives, à la linguistique...

Ce mémoire s'efforce de poursuivre les travaux déjà engagés dans l'un de ces domaines, celui qui analyse les données recueillies sous un angle linguistique, afin de dégager, si possible, des modèles et des constantes dans le dialogue homme/machine.

A travers cette analyse de données - formulations en langage naturel, termes entrés par les utilisateurs au cours de leur session de recherche - qui occupe une position centrale dans ce travail, nous espérons apporter des éléments de réponse aux questions suivantes :

Que peut apporter la linguistique à l'étude des systèmes d'information ? Nous tenterons de situer et de mesurer la place des études linguistiques dans la recherche sur les catalogues en ligne, en nous référant à la littérature consacrée au traitement linguistique dans les OPACs et les bases de données.

Comment peut-on envisager de poursuivre l'enquête PARINFO-OPACs et que peuvent attendre bibliothèques et usagers de ces prolongements ? La troisième partie de ce travail formulera quelques propositions pour l'élaboration d'un nouveau protocole expérimental.

Notre étude s'appuie sur une partie des sessions enregistrées en Juin 92, à la Médiathèque de La Villette, dans le cadre du projet PARINFO-OPACs. Ce contexte particulier appelle quelques remarques préalables.

M. Mohamed Hassoun, Maître de Conférences à l'ENSSIB, a assuré le traitement informatique des données recueillies durant l'enquête : il a notamment élaboré des fichiers de synthèse, un pour chaque session, qui organisent l'essentiel de ces données (questions en langue naturelle du préquestionnaire et du postquestionnaire électroniques, tableau des termes et des commandes entrés par l'utilisateur...). Nous avons d'ailleurs constitué notre corpus en nous appuyant sur le dénombrement des références affichées tel qu'il est fourni par la ligne "nb de réf. affichées" de chaque synthèse.

Il n'était pas envisageable de reproduire l'intégralité des sessions étudiées : le document ainsi obtenu (entre 800 et 1000 pages) n'aurait pas été consultable. L'ensemble des données occupe 8 disquettes 3 pouces 1/2, 10 si on prend en compte les questionnaires électroniques, de sorte qu'il ne nous a pas été possible non plus de les joindre sous cette forme. Nous avons dû nous résoudre à ne fournir que les données indispensables, extraites des fichiers de synthèse et réunies dans un volume annexe. L'ensemble des données recueillies à la Médiathèque de La Villette est conservée à l'ENSSIB, sous forme informatique.

Le lecteur trouvera deux types de listes dans l'annexe 2. La première reproduit le classement par nombre d'étapes et mode(s) de recherche que nous adoptons dans le second chapitre :

- Sessions en une étape et un mode de recherche (46)
- Sessions en plusieurs étapes et un mode de recherche (39)
- Sessions en plusieurs étapes et deux modes de recherche (31)
- Sessions en plusieurs étapes et 3/4 modes de recherche (6)
- Sessions présentant des ambiguïtés (6)
- Sessions en plusieurs recherches indépendantes (26)

La seconde partie de l'annexe présente, pour chaque session, la réponse aux questions 24 du préquestionnaire et 11 du postquestionnaire, ainsi que les modes de recherche choisis par l'utilisateur (Titre, Sujet, ...) et les termes entrés durant sa recherche. Les sessions sont présentées dans un ordre alphabétique et numérique, qui permet au lecteur de se reporter directement à une session donnée (ordre alphabétique des postes, puis ordre chronologique des jours d'enregistrement, puis, si nécessaire, numéro d'ordre croissant de la session sur ce poste et pour ce jour).

Citer ces divers enregistrements au cours de l'étude nécessite enfin d'utiliser un certain nombre de codes, notamment pour distinguer chaque session. Afin de faciliter la lecture des résultats, l'annexe 4 du second volume propose une liste de ces codes et abréviations.

CHAPITRE 1

Traitement de la requête en langue naturelle

Le traitement linguistique dans les OPACs et les bases de données intervient à deux "moments" essentiels de la conception et de l'utilisation des systèmes : la structuration des données de la base et l'interprétation des requêtes. Nous nous intéresserons plus particulièrement à cette seconde étape et tenterons de dégager des orientations dans l'abondante littérature consacrée aux systèmes de recherche d'information: comment peut-on envisager le traitement de la question de l'utilisateur, notamment lorsque celle-ci est formulée en langage naturel ? Par commodité, nous distinguerons cette étape "traitement de la requête" de l'étape "structuration des données", bien qu'elles soient en pratique difficilement dissociables.

Nous mettrons notamment l'accent sur les perspectives actuelles, qui associent l'étude et le traitement linguistiques à d'autres techniques et disciplines pour améliorer la convivialité des systèmes et prendre en compte le langage et le comportement de l'utilisateur.

Nous nous intéresserons alors à l'un des projets récemment lancés dans ce domaine, le projet PARINFO-OPACs, recherche exploratoire sur les OPACs menée dans le cadre du programme PARINFO, et qui comporte un important volet linguistique.

I- Le traitement linguistique de la requête

Le traitement linguistique revêt une importance capitale pour la recherche des documents et des informations. Les systèmes documentaires ont pour objectif de répondre à des besoins de natures diverses. H. Le Crosnier en a dressé la typologie suivante¹ :

- *besoins de vérification*, qui donnent lieu à une recherche ciblée
- *besoins conscients concernant un sujet*, qui se traduiront par un passage en revue des documents relatifs à un sujet précis et connu.
- *besoins flous concernant un sujet*, les plus difficiles à cerner et à formaliser pour l'utilisateur, qui souhaite explorer de nouveaux concepts sur des sujets non connus.

On appelle alors "requête" la question telle que l'utilisateur la fournit au système et telle que celui-ci doit la traiter. Le rôle du système consiste à établir une correspondance entre le ou les thèmes contenus dans la requête et le contenu informationnel des documents de la base.

Les catalogues et autres systèmes de recherche documentaire ont longtemps contraint l'utilisateur à formuler seul sa question, en utilisant le langage spécifique d'interrogation du système. C'est encore le cas de la plupart des OPACs à disposition dans les bibliothèques, dont celui de la Médiathèque de La Villette. Pourtant, de nombreuses études montrent que l'utilisateur non spécialiste, qui ne connaît ni la base documentaire, ni l'indexation des documents, rencontre de grandes difficultés pour mener à bien son interrogation, surtout dans le cas d'une recherche par sujet.

1- LE CROSNIER, H. **Systèmes d'accès à des ressources documentaires : vers des antéserveurs intelligents** : Thèse. Marseille : Université de droit, d'économie et des sciences d'Aix-Marseille III, Faculté des sciences et techniques de Saint-Jérôme, 1990, p. 47.

Ces découvertes ont donné lieu à la réalisation d'interfaces et de systèmes évolués, destinés à prendre en charge le traitement linguistique de la question et sa traduction dans le langage interne du système. On assiste actuellement au développement de systèmes experts qui devraient remplacer le documentaliste auprès de l'utilisateur.

1- Typologie des catalogues en ligne

Selon C. Hildreth, on peut construire une typologie des catalogues en ligne en distinguant trois générations de catalogues. Actuellement, la plupart des OPACs appartiennent à la deuxième génération, tandis que les OPACs de la troisième génération restent à l'état de prototypes expérimentaux. Ces trois générations correspondent à des stades différents d'évolution des techniques de recherche assistées par ordinateur et des interfaces homme-machine. On y remarque en particulier une complexification du traitement linguistique, sur le modèle des bases de données documentaires :

Les OPACs sont nés dans les années 60, dans les bibliothèques anglaises et américaines, de l'automatisation de deux grandes fonctions, la gestion du prêt et le catalogage partagé en ligne. Les premiers catalogues, qui suivaient les principes de la précoordination, n'étaient que des versions simplifiées des catalogues traditionnels sur papier ou microfiches.

Les catalogues "de seconde génération" sont issus d'un rapprochement, pas toujours homogène, avec les systèmes commerciaux d'interrogation de bases de données. Ces derniers, dès les années 60, ont utilisé la méthode de recherche d'information par "mot libre" ou postcoordination et la méthode booléenne (cf. infra § 2-2).

Il faut cependant distinguer catalogues en ligne et bases de données bibliographiques. Les enregistrements d'un catalogue en ligne sont relativement dépourvus de texte, leur contenu est souvent moins spécialisé que celui d'une base de données et leur indexation par sujet moins étendue. N. Mitev et C. Hildreth concluent que *les catalogues interactifs de deuxième génération peuvent être considérés comme des systèmes de recherche d'information à part entière, mais d'une nature particulière*².

Les catalogues de troisième génération prennent en compte certaines innovations concernant notamment le langage et le comportement des utilisateurs. Nous reviendrons ultérieurement sur les aspects linguistiques de ces améliorations.

2- MITEV, N., HILDRETH, C. Les catalogues interactifs en Grande-Bretagne et aux Etats-Unis : Systèmes et interfaces. *Bulletin des Bibliothèques de France*, 1989, Vol. 34, n° 1, p. 25.

2- Modèles de recherche documentaire

2-1 Différents modèles de recherche documentaire

Tout système de recherche d'information s'appuie sur une structure qui *définit le modèle de documents, ainsi que la méthodologie d'évaluation de la correspondance entre requête et documents*³. J. Nie distingue différents modèles, plus ou moins répandus et évolués : booléen, vectoriel, probabiliste, logique, sémantico-linguistique, tous développés durant les années 70. Il note cependant que l'on s'oriente actuellement vers la définition de modèles plus généraux.

Les OPACs et bases de données s'appuient principalement sur le modèle booléen.

2-2 Le modèle booléen et ses limites

2-2-1 Le modèle booléen standard

Dans ce modèle, un document est représenté par un ou plusieurs mots-clés. Une requête est une expression, elle aussi composée de mots-clés ou de descripteurs libres ou contrôlés, (selon le système documentaire), connectés par des opérateurs logiques : ET, OU, SAUF.

G. Van Slype a formalisé les étapes d'une recherche dans un tel système⁴. Certaines, les suivantes, impliquent un traitement linguistique :

- formulation mentale de la requête en langage naturel.
- traduction des concepts de la question dans le vocabulaire du système documentaire interrogé, par l'utilisateur lui-même ou par un intermédiaire spécialiste.
- mise en équation de la question par utilisation des relations syntaxiques autorisées par le système pour lier les mots-clés ou les descripteurs : ce sont ici les opérateurs booléens.
- éventuellement, extension de l'équation de recherche.
- jugement de pertinence sur les documents extraits.

Certaines techniques ont été développées afin de préciser ou d'étendre les questions.

Les troncatures permettent d'élargir une question en ne tenant pas compte des suffixes, des préfixes ou de groupes de caractères à l'intérieur des mots-clés. Cette méthode, bien que non linguistique, permet d'éviter de mentionner les variations flexionnelles d'un mot, ce qui s'avère utile dans un système fondé sur une indexation libre.

Grâce aux opérateurs de proximité, on peut exercer une contrainte sur la situation respective des descripteurs : exiger que deux descripteurs soient adjacents ou situés à n mot(s) au plus l'un de l'autre. Cette méthode, tout en présentant aussi un risque de bruit et d'élimination de documents

3- NIE, J. **Un modèle logique général pour les Systèmes de Recherche d'Informations : application au prototype RIME** : Thèse. Grenoble : Université Joseph Fourier Grenoble 1, Laboratoire de Génie Informatique IMAG, 1984, p. 4.

4- VAN SLYPE, G. **Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires**. Paris : Les Editions d'Organisation, 1987, p. 222.

pertinents, constitue un réel progrès par rapport aux opérateurs booléens, qui n'exercent pour leur part aucune contrainte sur la situation des descripteurs dans le document.

Certains systèmes exploitent en outre les relations sémantiques d'un thésaurus, ce qui leur permet d'ajouter automatiquement les descripteurs équivalents, génériques/ spécifiques, ou associés au descripteur entré. Ce sont, par exemple, les systèmes STAIRS-TLS, MISTRAL et GOLEM⁵.

2-2-2 Limites du modèle booléen et des systèmes de recherche "classiques"

Depuis la question posée par C. Hildreth en 1983 : "**To boolean or not to boolean**", de nombreux articles et études s'accordent à dénoncer les limites de la logique booléenne pour l'utilisateur non averti. N. Mitev et C. Hildreth le soulignent⁶:

Ces techniques sont bien adaptées pour les bases de données spécialisées, indexées en profondeur et qui comportent de longs abstracts, quand l'interrogation est faite par des intermédiaires professionnels ou des utilisateurs finals formés [...] expérimentés et assidus - et encore ces derniers peuvent eux aussi avoir des problèmes.

Cette critique s'adresse aussi bien aux OPACs qu'aux bases de données installées sur des serveurs commerciaux.

Sur le plan linguistique, on peut déceler divers problèmes ou lacunes, dont certains perturbent également les recherches non booléennes :

Des problèmes morpho-lexicaux

Les utilisateurs de catalogues ne sont pas formés à l'usage des vocabulaires contrôlés d'indexation par sujet. Ils ne connaissent ni le vocabulaire, ni la technique d'indexation, ni la structure des fichiers d'un système ou d'un établissement donné. Ils éprouvent donc de réelles difficultés à faire correspondre leurs propres termes ou formulations d'un concept avec le langage de la base. Lorsque l'utilisateur retrouve les mots retenus pour l'indexation, ceux-ci peuvent subir des déformations : fautes d'orthographe, variantes graphiques, flexionnelles, dérivation ou synonymie.

N. Mitev et C. Hildreth posent un problème linguistique capital : comment l'utilisateur pourrait-il formuler une question sur un sujet qui, précisément, lui est encore inconnu ou méconnu ? L'utilisateur interroge un système de recherche documentaire parce qu'il a décelé un manque dans son savoir : exprimer sa requête ne peut que mettre en évidence les problèmes conceptuels liés à ce manque.

Problèmes syntaxiques

Le mécanisme de recherche booléenne est rigide : l'utilisateur doit retrouver le ou les termes associés au document recherché, mais aussi la manière exacte de les apparier. Or, il maîtrise souvent mal les différentes techniques de recherche et de reformulation disponibles.

5- VAN SLYPE, Id., p. 87.

6- MITEV, N., HILDRETH, C., Id., p. 25.

Les opérateurs booléens opèrent une simple mise en relation de descripteurs, sans prendre en compte la nature du lien qui les unit. Ils ne permettent aucune structuration des données de la base. Cette lacune fondamentale a orienté un certain nombre de travaux linguistiques qui visent à rapprocher les descripteurs en se fondant sur des critères formels.

2-3 Un OPAC de la seconde génération : l'OPAC de la Médiathèque de La Villette

La Médiathèque de la Cité des Sciences et de l'Industrie de La Villette dispose d'un catalogue informatisé depuis 1985. En 1990, la première version GEAC 8000 a été remplacée par la version GLIS 9000. Il s'agit d'un OPAC de la seconde génération.

Les accès au catalogue sont au nombre de six :

- TIT : recherche par titre
- AUT : recherche par auteur
- A-T : recherche par auteur-titre
- SUJ : recherche par sujet
- NUM : recherche par numéro (cote, indice, éditeur, ISBN, ISSN, etc.)
- MOT : recherche par mot(s) du titre, de l'auteur et/ou du sujet.

L'accès MOT est le seul qui permet de conduire une recherche booléenne, proposée au lecteur selon deux modes différents :

- ACC = recherche mot accélérée (l'utilisateur élabore sa clé de recherche à partir d'un simple mode d'emploi et de quelques exemples).

- CPX = pour un mode d'emploi (CPX offre une assistance à la recherche).
L'enquête réalisée sur l'OPAC de la Médiathèque a mis en évidence la faible utilisation de l'accès par mot(s) : 7 à 10 % des recherches seulement. Notre étude des sessions conduites par les utilisateurs permet également de constater que la plupart d'entre eux choisissent le mode ACC et conduisent donc leur recherche sans assistance. Il s'ensuit des erreurs de syntaxe et des refus du système qui découragent fréquemment les utilisateurs. L'attitude la plus fréquente est alors de ne pas combiner les termes et de proposer au système un seul mot du titre, de l'auteur ou du sujet : autrement dit, l'utilisateur renonce souvent à mener une recherche booléenne, ce qui confirme les difficultés évoquées précédemment.

S'appuyant sur ces constatations, les recherches en cours et les premières réalisations ont tenté de prendre en charge les problèmes linguistiques de l'utilisateur, en lui offrant notamment de formuler sa requête en langue naturelle.

3- Les systèmes s'attachant au traitement de la question

On peut construire une typologie des systèmes qui traitent le problème de la question, avec G.P. Zarri⁷:

- les interfaces évoluées : ce sont les outils qui ont pour objectif de servir d'intermédiaires entre l'utilisateur et un système documentaire "classique". Certaines permettent la formulation de la question en langage naturel.
- les systèmes reposant sur une analyse linguistique de la base documentaire, dont certains traitent de manière identique les documents et les questions des utilisateurs.
- les systèmes experts de recherche documentaire, qui constituent des intermédiaires linguistiques et documentaires entre une base et un utilisateur.

3-1 Les interfaces

3-1-1 Quelques réalisations

Cette technique est désormais répandue, notamment dans le domaine industriel (On pourra consulter certains articles réunis par GSI-ERLI⁸, qui présentent un certain nombre d'applications mises en place par cette société). Certaines interfaces ont pour but premier d'apporter une assistance de type documentaire à l'utilisateur. D'autres, auxquelles nous nous intéresserons plus particulièrement, lui offrent d'abord un soutien linguistique. Selon R. Dachelet, les interfaces en langage naturel ont d'abord eu pour objectif de *projeter l'utilisateur, ou plutôt sa requête, sur la base ou les bases*. Puis elles se sont dotées des moyens d'interpréter le sens de la question et de construire des stratégies de recherche⁹.

Certaines interfaces ne traitent pas le langage naturel, mais un langage restreint : ainsi le projet **CITE**, élaboré à la National Library of Medicine du Maryland (USA), qui permet de rechercher des informations dans les deux grandes bases documentaires de littérature médicale MEDLINE et CATLINE.

Pourtant, les interfaces en langage naturel avec les bases de données commencent à se développer¹⁰.

Le système **IRUS** (Intelligent Retrieval Using the RUS Parsing System), a été développé par BBN (Bolt, Beranek and Newman), à Cambridge (USA). Ce système utilise l'analyseur syntaxique du langage naturel RUS. Associé à un interpréteur sémantique, cet analyseur permet de produire une

7- ZARRI, G.P. Etat de l'art : les nouvelles tendances de l'informatique documentaire. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1988, n°32, p.11-40.

8- GSI-ERLI. **Ingenierie linguistique et documentaire : Recueil d'articles**. Charenton-le-Pont : GSI-ERLI, Mai 1991. 166 p.

9- DACHELET, R. Etat de l'art de la recherche en informatique documentaire : la représentation des documents et l'accès à l'information. In INRIA. **Le document électronique, Cours INRIA, 11-15 Juin 1990**. Rocquencourt : INRIA, 1990, p. 118.

10- NORMIER, B. Interfaces en langage naturel avec des bases de données. **Télécom**, n°87, 1991, p. 36-39.

représentation de la question dans un langage de représentation des connaissances (MRL = Meaning Representation Language).

L'interpréteur utilise un dictionnaire et des règles d'interprétation sémantique, propres à un domaine d'application particulier. Le système peut donc être transporté à un autre domaine ou à plusieurs bases. Un module utilise la représentation de la question pour piloter les commandes du système documentaire auquel l'interface est couplée.

D'autres interfaces sont mises au point dans le domaine industriel. On citera notamment le **Natural Language System** (NLS), interface en langage naturel destinée à l'interrogation des pages jaunes et du guide des services Minitel¹¹. Le système proposé constitue une alternative à la technique de recherche traditionnelle par des mots-clés pointant sur les rubriques de la nomenclature. Le NLS facilite l'accès aux rubriques, par l'association d'une base de connaissances spécifique à la liste des rubriques et par la construction d'une grammaire assurant l'interprétation des requêtes.

Le traitement de la requête s'effectue en plusieurs étapes:

- l'automate d'analyse syntaxique utilise d'abord un système d'analyse lexicale et morphologique pour la reconnaissance des mots entrés par l'utilisateur. A ce stade sont traités les problèmes d'accentuation, les problèmes d'orthographe, les mots composés... L'automate dispose d'un dictionnaire du français courant.
 - un système d'analyse syntaxique de la phrase reconnaît la "tête" de phrase et assigne des "fonctions de base" à chacun des composants.
 - un système d'analyse sémantique recherche le ou les index capable(s) de répondre à la question.
- L'interface ne modifie pas l'architecture générale du système, mais l'utilisateur n'a plus à rechercher lui-même les rubriques susceptibles de l'intéresser¹².

3-1-2 Critique de ces systèmes

Une branche de la recherche sur les systèmes d'information documentaire, représentée notamment par J-C Bassano, critique cependant le principe des interfaces¹³ :

Il ne peut s'agir désormais d'une "interface intelligente et relativement mécanique" installée devant un système classique pour la recherche de l'information documentaire. Ces systèmes - conçus dans les années 1980 - assistent et aident les utilisateurs lors de leur formulation initiale : [...] respect des règles syntaxiques et/ou transcodification vers des langages d'interrogation spécifiques.

J-C Bassano leur reproche de ne pas prendre en charge l'ensemble de la stratégie de recherche, mais de se contenter d'effectuer *quelques opérations (mécaniques) simples.*

11- CLEMENCIN, G. Querying the French Yellow Pages : natural access to the directory. **Information processing & management**, 1988, Vol. 24, n°6, p. 633-649.

12- Id., p. 634.

13- BASSANO, J-C. Systèmes experts et stratégies de recherche de l'information. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1988, N°32, p. 50.

Il leur oppose des systèmes documentaires prototypes capables de reproduire les outils et les techniques des documentalistes experts : les *systèmes experts pour la recherche documentaire*.

3-2 Les systèmes traitant de la même manière documents et questions

Un certain nombre de travaux reposent sur l'idée que les documents textuels d'une base documentaire donnée et les questions des utilisateurs formulées en langage naturel fournissent un même type de représentation. A ce titre, il convient de leur appliquer le même traitement, afin d'en extraire et d'en comparer les éléments porteurs d'informations. Le processus de recherche repose alors sur une double analyse linguistique, parallèle, des documents et des questions, puis sur un appariement des éléments informationnels extraits.

On retrouve ce type d'approche dans les travaux du CRISS de Grenoble, qui appliquent le modèle linguistique d'extraction des syntagmes nominaux défini par A. Berrendonner et M. Le Guern¹⁴. Après extraction des syntagmes nominaux de la requête, on procède à une comparaison avec les syntagmes nominaux d'un dictionnaire créé lors de l'indexation des documents. On part du syntagme nominal de base pour remonter ensuite dans les noeuds du graphe d'indexation, grâce aux autres syntagmes de la requête. Le système fournit à l'utilisateur la liste des documents associés aux noeuds pertinents.

SPIRIT offre, quant à lui, plusieurs formules d'interrogation :

- par une question booléenne
- par une question en langage naturel
- par une question sous forme d'un document de la base.

Rappelons que les travaux sur le système SPIRIT, Système Syntaxique et Probabiliste d'Indexation et de Recherche d'Information Textuelle, sont menés par l'équipe de MM. Andreewsky, Debili et Fluhr depuis le début des années 70. P. Radasoa¹⁵ définit ainsi le système dans sa thèse, qui a également contribué à le perfectionner :

Le système SPIRIT, opérationnel depuis 1981, est un système de recherche documentaire qui permet la génération automatique de bases de données textuelles sans préparation préalable des textes et leur interrogation en langage naturel. La réponse du système est une liste de documents triés selon leur proximité sémantique par rapport à la question posée. Le système utilise un traitement linguistique et un traitement statistique.

La question en langage naturel subit le même traitement que les documents de la base : un traitement linguistique. Puis le système recherche les documents ayant des termes communs avec la question. L'appariement entre les documents pertinents et la requête repose aussi sur des critères linguistiques, doublés de critères statistiques. Le système

14- Pour les fondements théoriques et les applications de ce modèle, on pourra notamment consulter les travaux d'A. BERRENDONNER, M. LE GUERN, R. BOUCHE, J.P. METZGER, M. DE BRITO, J. ROUAULT.

15- RADASOA, H.P. **Méthode d'amélioration de la pertinence des réponses dans un système de bases de données textuelles** : Thèse. Paris : Université de Paris Sud (Orsay), 1988. 156 p.

calcule une "proximité sémantique" entre les deux et hiérarchise les documents fournis à l'utilisateur sur une échelle de pertinence.

L'interrogation en langage naturel permet d'utiliser plusieurs fonctions¹⁶:

- une fonction qui engage une recherche sur les titres et les résumés des documents : QUES
- une fonction qui recherche sur toutes les zones du document, y-compris les champs structurés : SPQUES
- une fonction qui complète la question en l'analysant plus finement : SUITEQ.
- la fonction QUEDOC permet enfin de remplacer une question en langage naturel par un document entier, jugé pertinent par l'utilisateur. Ce texte sera alors traité comme une question textuelle. Le système recherche dans la base les documents qui ont le plus de mots en commun avec le document-question.

3-3 Les systèmes experts de recherche documentaire

D'après Chris Paice, cité par J-C. Bassano, ces systèmes experts présentent, sur le plan linguistique, deux caractéristiques essentielles :

- Ils acceptent comme questions des fragments de textes en langue naturelle.
- Ils enrichissent automatiquement ces requêtes et les transforment en une formulation plus efficace pour interroger la base.

Le prototype même de système expert correspondant à cette définition est le système **DIALECT**, élaboré par J-C. Bassano dans le cadre du CNRS et de l'Université de Dijon¹⁷. Ce système adopte pour point de départ la question telle que l'utilisateur est capable de la formuler en langage naturel : sa "requête naïve". Elle peut s'appliquer à un domaine de connaissances encyclopédique. L'analyse linguistique de la question se déroule en trois temps :

- une analyse morpho-lexicale
- la résolution des ambiguïtés grammaticales
- une analyse syntaxique.

A l'issue de ce traitement, la question de départ est constituée de zones de texte pertinentes ou "énoncés" : un énoncé se compose de deux descripteurs et des relations syntaxiques qui les unissent. La séquence d'énoncés génère alors une équation booléenne, qui intègre aussi tous les descripteurs que le système a jugé sémantiquement équivalents à ceux de la question.

Cette équation vise à extraire non pas des documents, mais les phrases des documents qui répondent à la question. La spécificité du système réside alors dans ce qu'il procède à un remodelage linguistique de la question. Il repose sur la constitution automatique d'une matrice de paraphrases, grâce à des procédures d'analyse distributionnelle permettant de repérer des régularités syntaxiques formelles. Ces procédures sont

16- IZABEL, Y. **Sur le traitement de l'information en langage naturel : le système SPIRIT**. Mémoire. Paris : Conservatoire National des Arts et Métiers, Institut National des Techniques de la Documentation, 1984. 90 f.

17- BASSANO, J-C. **DIALECT : un système expert pour la recherche documentaire** : Thèse Université Paris Sud. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1986, N°23 et 24.

pilotées par le noyau "système expert" de DIALECT. Il applique de nouveau la phase d'analyse linguistique pour dégager d'autres énoncés des phrases retenues et les intégrer dans des "structures élémentaires".

Des règles comparent les énoncés de départ aux structures élémentaires et ajoutent à la question les nouveaux énoncés pertinents. Le processus est réitéré jusqu'à une condition d'arrêt.

Dans le système IOTA¹⁸, le traitement initial de la requête suit les mêmes étapes que dans les systèmes "classiques" présentés au paragraphe 3-2:

- traitement automatique de la question, qui consiste à localiser les syntagmes nominaux pour élaborer une expression booléenne. Ce traitement est identique à celui des documents de la base.
- appariement entre syntagmes nominaux de la requête et termes retenus pour l'indexation.

Cependant, l'interprétation de la requête est confiée à un système expert qui simule l'activité d'un documentaliste. Ce système utilise un thésaurus, dans lequel les relations sont pondérées : il évalue la proximité entre chaque couple de mots en relation.

- Ce système évalue le niveau de l'utilisateur : par comparaison entre les termes de la requête initiale et ceux du thésaurus. Il calcule le "degré de dégradation" de la requête.
- Il interprète la requête finale élaborée, ce qui permet de fournir des références pondérées.
- Il remodèle automatiquement les requêtes.

Un tel système - mais il en est de même dans le système I3R¹⁹ - tire parti des recherches cognitives, et non plus seulement linguistiques. Il témoigne aussi d'une tendance forte de la recherche en informatique documentaire : les chercheurs s'intéressent davantage, désormais, au processus de recherche d'information, par opposition notamment à la représentation des documents ; cela les conduit nécessairement à étudier le point de vue de l'utilisateur et son langage afin de pouvoir les modéliser. D'où une nouvelle approche du traitement linguistique dans les OPACs et les bases de données.

18- DEFUDE, B. *Etude et réalisation d'un système intelligent de recherche d'information : le prototype IOTA* : Thèse de doctorat. Grenoble : USMG, 1986.

19- CROFT, W.B., THOMPSON, R.H. I3R : A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science*, 1987, Vol. 36, n°6, p. 389-404.

II- Le traitement linguistique au service de l'utilisateur

Toutes les recherches actuelles sur les OPACs, les bases de données et les systèmes experts tentent de répondre aux enquêtes réalisées sur le comportement des utilisateurs, qui mettent en évidence leurs difficultés devant les systèmes proposés (on citera pour mémoire les travaux de C. Borgman, de M. Hancock-Beaulieu...).

Le traitement linguistique évolue, dans deux directions complémentaires, mais souvent disjointes :

- celle qui crée ou perfectionne des outils linguistiques que les systèmes mettront à la disposition de l'utilisateur. Elles travaillent "sur l'existant".
- celle qui étudie le langage de l'utilisateur et ses "modèles mentaux", afin de modifier la conception même des systèmes de recherche d'information. La linguistique soutient ici les recherches en sciences cognitives.

1- Linguistique et convivialité des systèmes

On peut distinguer trois orientations de recherche :

- certaines études "confrontent" le vocabulaire initial de l'utilisateur à celui de l'OPAC ou de la base de données.
- d'autres visent à orienter l'utilisateur dans le système, afin qu'il puisse formuler ou enrichir sa question.
- d'autres encore prennent en charge les opérations de formulation et/ou de reformulation, ou aident le non-spécialiste à les réussir.

1-1 Etudes sur les langages documentaires : distance et niveau linguistiques

Une direction particulière du traitement linguistique dans les OPACs et les bases de données consiste à étudier la structuration des langages documentaires ou classificatoires, leur emploi par l'utilisateur et leur adéquation au système dans lequel ils sont implantés²⁰.

Une expérience a été conduite à l'Université de Californie Los Angeles (UCLA) pour évaluer la correspondance lexicale et syntaxique existant entre les termes entrés par des utilisateurs du catalogue en ligne de la bibliothèque de l'UCLA et les vedettes LCSH (Library of Congress Subject Headings ou vedettes-matières de la Bibliothèque du Congrès, qui sont utilisées dans un grand nombre de bibliothèques universitaires américaines et dans leurs catalogues en ligne).

Il s'agissait d'évaluer la concordance entre le vocabulaire de l'utilisateur et les vedettes LCSH lors d'une interrogation du catalogue, afin de mesurer la performance des vedettes. L'enquête confronte 171 expressions d'utilisateurs avec la vedette simple ou multiple contenue dans les LCSH et définit leur degré de similitude sur le plan lexical et syntaxique :

- correspondance exacte

20- CARLYLE, A. Matching LCSH and user vocabulary in the library catalog. **Cataloging & Classification Quarterly**, 1989, Vol. 10, n° 1-2, p.37-63.

- correspondance exacte, à une variation de ponctuation près
 - correspondance exacte, mais variation de l'ordre des mots
- etc.

Elle ne tient pas compte des correspondances sémantiques, pourtant fondamentales, mais qui supposeraient une autre approche méthodologique. L'analyse lexicale et syntaxique devrait donc être complétée, dans l'avenir, par une analyse sémantique. Mais cette étude montre déjà comment la linguistique peut aider à améliorer le fonctionnement d'un système en ligne.

Elle permet en effet de tester le niveau linguistique du langage documentaire utilisé par le catalogue. Dans le cas des LCSH, les questions qui se posent sont les suivantes : les vedettes sont-elles trop spécialisées (elles privilégieraient le niveau spécifique aux dépens du générique) ou pas assez ? Sont-elles à jour ? Quel niveau de hiérarchisation le public utilise-t-il spontanément ?

Elle s'efforce aussi d'analyser le langage de l'utilisateur, pour ne plus lui "plaquer" un langage d'interrogation trop éloigné du sien. On tente de mesurer ici une distance linguistique entre système et utilisateur.

A. Carlyle conclut à la nécessité de développer certains thèmes de recherche, qui lient étude linguistique, comportement de l'utilisateur et "design" des catalogues en ligne :

- En quoi la structure et le contenu de l'interface d'un catalogue affectent-ils la structure et le contenu des expressions utilisées par l'utilisateur ?
- Quelle connexion sémantique existe-t-il entre ce que dit et ce que pense l'utilisateur d'un catalogue en ligne ?
- Comment améliorer la construction des vedettes LCSH, comment développer le langage d'un catalogue, pour accroître la satisfaction de l'utilisateur ?
- Un langage d'indexation convient-il mieux à un système en ligne qu'à un autre ?

Des expériences proches ont été conduites sur la Classification Décimale Dewey, mettant en évidence les possibilités d'enrichissement du vocabulaire d'une base et par suite des recherches effectuées.

1-2 Techniques de navigation dans les OPACs

Une première prise en compte des blocages des utilisateurs vis à vis des méthodes de recherche booléenne a conduit les concepteurs à équiper les OPACs et bases de données de procédures de butinage, de navigation (ou "browsing"). Elles permettent à l'utilisateur de feuilleter des listes alphabétiques lorsqu'il ne sait pas formuler sa question ou souhaite explorer des domaines peu familiers. Il s'agit souvent cependant d'une exploration linéaire, insuffisante.

Seuls quelques prototypes proposent des explorations non linéaires, d'un concept à un autre de la base, notamment le système COREL. Le système le plus évolué est, dans ce domaine, I3R : l'un de ses systèmes experts, le "browsing expert", permet à l'utilisateur de naviguer dans la base de connaissances en partant d'un document, d'un auteur, d'un descripteur, puis en passant à d'autres objets de la base.

Les systèmes "classiques" laissent encore le soin à l'utilisateur de formuler lui même sa requête et de l'enrichir. Certains prototypes dépassent cependant ce stade et réalisent cette opération : ils utilisent des techniques de reformulation.

1-3 La reformulation

La thèse de P. Radasoa est consacrée aux diverses techniques de reformulation envisageables, en l'occurrence dans le système **SPiRiT**. Elles s'appuient en grande partie sur un traitement linguistique.

L'importance du traitement linguistique et du traitement statistique dans un système de recherche documentaire est mise en évidence, surtout lorsqu'on effectue une reformulation²¹.

1-3-1 Définition et méthodes

L'objectif premier de la reformulation consiste à trouver un chemin pour apparier les notions contenues dans les documents et les questions. Il existe trois types de reformulation : celle des documents de la base, coûteuse, celle de la question, qui entraîne un gonflement excessif des termes de cette question, et la reformulation conjointe des documents et de la question.

Différentes méthodes de reformulation assistée peuvent être distinguées ; P. Radasoa en dresse une liste non exhaustive :

- la normalisation

Elle consiste à choisir un seul représentant pour des mots ou groupes de mots ayant le même sens, ou pour des mots appartenant à une même famille, ou encore pour désigner une notion (on utilise alors un thésaurus). Le système d'interrogation des pages jaunes de l'annuaire repose sur une normalisation sémantique.

- l'utilisation des mots de la même famille

On regroupe en famille des mots ayant le même radical, en conservant l'information portant sur leurs différences morphologiques. Le système peut ensuite, lorsque les relations sémantiques le permettent, substituer un mot à un autre pour enrichir la question ou la normaliser.

- le thésaurus

Il met les mots en relation sémantique, et non plus morphologique. (La base de données Dianeguide propose notamment à l'utilisateur un élargissement automatique de sa recherche par sujet qui repose sur l'exploitation d'un thésaurus : à partir des descripteurs isolés dans la question originale, le système produit une liste de descripteurs voisins dans ce thésaurus et effectue une nouvelle recherche, en classant les réponses obtenues en fonction de la distance sémantique qui sépare ces nouveaux descripteurs des descripteurs de départ.)

- le correcteur orthographique

Il nécessite une analyse morphologique, puis une analyse syntaxique. Il doit aussi traiter le problème particulier des noms propres.

- l'explicitation

Elle consiste à reconnaître les structures phraséologiques et à traduire, au niveau interne, certaines notions implicites en termes explicites : interpréter, par exemple, que la structure [on appelle + Nom Commun] introduit vraisemblablement une définition.

- l'apprentissage de règles de reformulation

Certains systèmes sont capables "d'apprendre" les reformulations jugées pertinentes par un utilisateur et de les ranger dans un thésaurus particulier, qui constitue alors un moyen de corriger les erreurs d'une classe d'utilisateurs donnée.

- l'interrogation par partie du document qui, rappelons-le, permet d'utiliser tout ou partie d'un document pertinent pour interroger la base.

Les prototypes de recherche documentaire combinent parfois plusieurs de ces outils linguistiques.

1-3-2 Quelques systèmes proposant une reformulation

La plupart des prototypes ont recours à la technique de "relevance feedback", une méthode qui associe statistique et linguistique. Elle consiste à utiliser les résultats d'une première recherche pour reformuler automatiquement la requête, en augmentant ou en diminuant les poids des descripteurs, selon que l'utilisateur a jugé le document pertinent ou non. Cette technique accroît de manière significative la précision des réponses.

Le système **DIALECT** génère cette opération automatiquement; elle est prise en charge par ses outils d'analyse linguistique, qui sélectionnent des zones de texte pertinentes jusqu'à l'obtention d'une condition d'arrêt.

Le système **SPIRIT** demande à l'utilisateur de sélectionner les documents ou parties de documents pertinents pour la recherche ; il les classe par degré de pertinence, puis il enrichit la question : son objectif est de diminuer le silence. C'est le noyau "système expert" qui gère la reformulation. Il s'appuie sur des bases de connaissances, groupes de règles correspondant à chacune des méthodes de reformulation possibles ; des méta-règles contrôlent le processus global de reformulation.

De même, dans le système **I3R**, c'est l'utilisateur qui juge de la pertinence des documents rappelés. Son évaluation est alors intégrée au système expert "request model builder", celui qui a modélisé la requête de l'utilisateur²².

Au contraire, le système **IOTA** pondère les descripteurs et commence sa recherche avec les mots les plus pertinents ou utilise ces mots pour enrichir la question.

Selon R. Dachelet, le développement de ces techniques est révélateur d'une évolution dans les buts que se fixe la recherche documentaire : *On a renoncé à l'espoir de fournir à l'utilisateur LA réponse à LA question posée. Une réponse, c'est aujourd'hui un document ou un ensemble de documents que l'utilisateur estimera pertinent(s)*²³.

22- CROFT, W.B., Id., p. 393.

23- DACHELET, R., Id., p. 126.

2- Langage de l'utilisateur et modèles mentaux

Les problèmes posés par la représentation des connaissances dans les systèmes de recherche d'information sont en fait commun à plusieurs sciences : sciences de l'information, Intelligence Artificielle, psychologie et linguistique, ainsi que le montrent les études menées par B. Vickery et son équipe²⁴.

Des travaux tels que ceux d'E. Kolmayer²⁵ tentent, à travers une expérimentation conduite sur le vocabulaire et les techniques d'interrogation d'un groupe d'utilisateurs, de dégager des modèles d'organisation des connaissances et de comprendre la nature de leurs blocages vis à vis de la recherche automatisée d'informations. En étudiant la variation des descripteurs choisis pour interroger sur des sujets imposés, on peut mettre en évidence les modèles - catégories logiques ou catégories écologiques : catégories naturelles, frames, scripts - auxquels fait appel un utilisateur. L'une de ses études a permis de constater, par exemple, que les usagers de bases de données ne savent pas déterminer à quel niveau hiérarchique (générique/spécifique) ils doivent interroger.

Un consensus réunit les chercheurs et concepteurs de systèmes autour de l'importance à accorder au traitement linguistique. Cette phrase de B. Menon nous paraît bien résumer la tendance actuelle :

Les textes, en tant qu'objets langagiers, posent avant tout - et qui s'en étonnera - des problèmes linguistiques. Il nous semble donc que le commun dénominateur des systèmes d'indexation automatique doit être une place centrale accordée à la résolution de ces problèmes.

Le traitement linguistique dans les OPACs et les bases de données a cependant connu deux évolutions majeures durant les dix dernières années.

Les études consacrées à l'indexation automatique et à l'interprétation des requêtes en langue naturelle ont montré que les traitements morpho-syntaxiques, de mieux en mieux maîtrisés, ne pouvaient suffire à résoudre tous les problèmes puisqu'ils ne permettaient pas de prendre en compte la signification des énoncés. De ce constat sont nées ce qu'Yves Chiaramella appelle *des stratégies de cohabitation entre des outils purement linguistiques (tels que des analyseurs syntaxiques) et des outils de manipulation de la connaissance*, qui font appel aux techniques de l'Intelligence Artificielle.

D'autre part, les systèmes à l'étude se doivent désormais de prendre en compte les difficultés de l'utilisateur. Le concept même de système de recherche d'information s'élargit pour intégrer l'utilisateur et instaurer un dialogue avec lui : ces travaux sur la convivialité des systèmes font également appel à la linguistique. Dans ce contexte, les tentatives de modélisation de l'utilisateur et de ses requêtes constituent à la fois une orientation pour la recherche et un enjeu pour le développement des OPACs et des bases de données grand public.

24- VICKERY, B.C. Knowledge representation : a brief review. *Journal of Documentation*, 1986, Vol. 42, n° 3, p. 145-159.

25- KOLMAYER, E. Bases de données grand public et organisation des connaissances. *Bulletin des Bibliothèques de France*, 1992, Vol. 37, n° 6, p. 8-13.

III- Une recherche exploratoire : le programme PARINFO-OPACs

1- Le programme PARINFO

Le Programme d'Aide à la Recherche en Information (PARINFO) a été lancé en 1990 par le ministère de la Recherche et de la Technologie et celui de l'Education nationale, de la Jeunesse et des Sports.

Partant du constat que le domaine de l'information spécialisée connaît actuellement d'importantes mutations, liées notamment à une évolution rapide des technologies, ces deux ministères ont mis en place, conjointement, un programme de financement de la recherche. Un appel à propositions a été lancé dans ce sens, destiné à amener des chercheurs de disciplines et d'horizons différents à collaborer sur cinq thèmes spécifiques²⁶ :

- modélisation et représentation des connaissances et des processus de raisonnement du point de vue des contenus informationnels, des flux et des accès
- méthodes et outils d'évaluation des systèmes d'information
- étude des démarches et pratiques informatiques des utilisateurs
- l'information dans le processus de compétitivité et d'innovation des entreprises
- l'information, ressource stratégique et industrie nouvelle.

Chaque proposition, examinée et retenue par un comité de sélection, devait être réalisée sur une durée maximum de deux ans. C'est dans ce cadre qu'a été élaboré un projet de recherche exploratoire sur les catalogues en ligne accessibles par le public, projet déposé en Juin 91 et développé en 92-93.

2- Le projet PARINFO-OPACs

2-1 Objectifs

Le projet est né d'une double prise de conscience.

- Le fonctionnement des catalogues en ligne est marqué par de graves insuffisances, que de nombreuses études ont contribué à mettre en évidence. Nous en citerons quelques-unes pour mémoire : celles de C. Borgman, de P. Le Loarer, de J. Le Marec. Il devenait donc urgent d'appeler des chercheurs appartenant aux différentes disciplines impliquées dans le sujet à coopérer et à mettre en place un programme de travail commun : conservateurs, informaticiens, psychologues, linguistes, chercheurs en sciences de l'information... Tels furent un des objectifs et une des conséquences du projet.

- Les chercheurs français sont encore peu présents dans ce champ de recherche fondamentale et appliquée que constituent les OPACs : il fallait donc ouvrir des perspectives de recherche en France et définir des thèmes de recherche pour l'avenir.

26- MINISTERE DE LA RECHERCHE ET DE LA TECHNOLOGIE, MINISTERE DE L'EDUCATION NATIONALE, DE LA JEUNESSE ET DES SPORTS. Programme d'aide à la recherche en information (PARINFO). *Brisés*, 1991, n°16, p.81-82.

La démarche adoptée par le groupe de projet se définit clairement dans cette phrase, extraite du rapport publié en 1993 sur le projet PARINFO-OPACs²⁷ :

L'idée est de structurer l'approche sur les OPACs et d'aboutir en sortie d'étude à une série de recommandations et de propositions utiles pour les bibliothèques (et leurs utilisateurs), le monde de la recherche en sciences de l'information en indiquant des pistes de développements et réalisations sur ce thème ou des thèmes très apparentés.

2-2 Les participants au projet

Conformément aux exigences du programme PARINFO et aux objectifs multidisciplinaires retenus pour cette recherche exploratoire, le projet a réuni plusieurs équipes, venues d'horizons différents :

- une équipe composée de chercheurs de l'ENSSIB²⁸, sous la direction de Danielle Roger, Conservateur de bibliothèque. Cette équipe de recherche en Sciences de l'information a pris en charge le pilotage du projet, le traitement informatique et statistique des données recueillies et la direction de recherche relative aux stratégies des utilisateurs d'OPACs.
- une équipe de recherche britannique, dirigée par Micheline Hancock-Beaulieu, de la City University, Department of Information Science, Centre for Interactive Systems Research, de Londres. Elle a apporté son soutien méthodologique et son expérience considérable dans le domaine des OPACs et de leur évaluation. (On doit à cette équipe un prototype de catalogue en ligne de troisième génération, OKAPI, qui s'appuie sur une expertise pour traiter les requêtes de l'utilisateur).
- une société spécialisée en ingénierie linguistique et documentaire, GSI-ERLI : elle possède une expérience importante en matière de langages documentaires, d'indexation automatique et a développé un savoir-faire dans le domaine des interfaces en langage naturel. (Elle a en particulier réalisé des interfaces d'accès aux pages jaunes pour France Télécom, Saritel et US-West). Cette société a plus particulièrement pris en charge l'exploitation linguistique des données recueillies au cours de l'enquête.
- la Médiathèque de la Cité des Sciences et de l'Industrie de la Villette a offert un milieu d'enquête "naturel", son OPAC et son public. L'équipe d'enquêteurs a été coordonnée par Maria Witt, Conservateur à la Médiathèque. La Médiathèque, qui a joué un rôle de pionnière en installant le premier OPAC en France, manifeste depuis longtemps son intérêt pour les questions relatives aux catalogues en ligne : elle a déjà organisé une enquête test sur l'OPAC auprès de ses abonnés, en 1990²⁹, et des journées d'étude sur ce thème.

27- ECOLE NATIONALE SUPERIEURE DES SCIENCES DE L'INFORMATION ET DES BIBLIOTHEQUES. CENTRE D'ETUDES ET DE RECHERCHES EN SCIENCES DE L'INFORMATION, et al., **Les catalogues en ligne accessibles par le public (OPACs) : recherche exploratoire**, p. 2 (cité plus loin comme [ECOLE 93])

28- Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques

29- BELET, F., **L'usage du catalogue informatisé : l'exemple de la Médiathèque de La Villette**, p. 22-31.

2-3 Terrain d'enquête

La Médiathèque est une bibliothèque multimédia, dotée d'un certain nombre de spécificités. Elle se compose de deux sections, situées dans des locaux proches mais distincts de la Cité des Sciences et de l'Industrie :

- la Médiathèque publique, qui propose au public un fonds à dominante scientifique et technique, en libre accès.
- la Médiathèque d'Histoire des Sciences qui réunit des documents sur l'histoire des sciences, des techniques, de la médecine, la muséologie des sciences et un fonds ancien de sciences. Son accès est réservé aux chercheurs et étudiants, ainsi qu'aux utilisateurs munis d'un laissez-passer.

Le catalogue en ligne est commun aux deux sections de la Médiathèque. Il offre en outre la possibilité de rechercher des documents dans une seule des deux sections, ou encore dans le fonds pour enfants exclusivement (grâce aux commandes LIM, puis LOC et à la sélection de MEDPUB, MEDENF ou MEDSPE).

Il s'agit donc d'un milieu d'étude bien défini, d'un public particulier et d'un fonds spécifique, à la fois grand public et spécialisé dans le domaine des sciences.

2-4 Méthodologie

2-4-1 Trois phases

Le travail réalisé comporte trois phases distinctes, mais complémentaires :

- tout d'abord une revue systématique et critique de la littérature consacrée aux catalogues en ligne, visant surtout à mettre en perspective des techniques élaborées dans d'autres disciplines, mais qui pourraient se voir transposées aux OPACs.
- la constitution et l'analyse d'un corpus de questions posées par les utilisateurs de l'OPAC de la Médiathèque de La Villette. L'enquête proprement dite a été précédée de 7 réunions du comité de pilotage, consacrées à la mise au point d'une méthodologie d'enquête - et notamment des questionnaires - puis à l'exploitation des résultats.
- une synthèse des différents travaux, qui a donné lieu à l'élaboration de thèmes de recherche. Résultats et thèmes sont consignés dans le rapport d'étude déjà cité³⁰.

Nous ne reviendrons pas sur la première phase du projet, dont on pourra trouver lecture dans ce même rapport. Il nous semble en revanche nécessaire d'évoquer les phases 2 et 3 du projet, afin de situer notre travail par rapport à l'enquête réalisée et aux directions définies.

2-4-2 Mise en oeuvre de l'enquête

Après une période de test et une pré-enquête (en Mai 92), qui ont permis d'ajuster les questionnaires, la collecte des données s'est déroulée du 3 au 28 Juin 1992, sur 5 postes de consultation de l'OPAC. Ces postes,

spécialement mis en place pour la durée de l'enquête, se situaient en divers points de la Médiathèque : postes ACC (Accueil), ME (MEIJ), MHS, TTA, UG (UGO).

Le corpus de données consiste en un enregistrement de la session de l'utilisateur, mais comprend également un préquestionnaire et un postquestionnaire électroniques, dont on trouvera un exemplaire en annexe 3 de ce mémoire. L'enquête électronique a été réalisée avec le logiciel OLIVE (Online Interactive Validation and Evaluation), que l'on a adapté au logiciel GEAC. Le tout a parfois été suivi d'un questionnaire oral, dirigé par un enquêteur. Le corpus final réunit 650 sessions, parmi lesquelles 598 s'avèrent finalement exploitables.

Divers traitements automatiques ont été appliqués à ces données de base, en particulier des traitements informatiques et statistiques, grâce aux logiciels Modalisa et Excel.

2-4-3 Perspectives

L'enquête réalisée en Juin 92 a d'ores et déjà été exploitée dans trois directions majeures :

- une étude de type statistique, qui a permis de réaliser une typologie des enquêtés³¹.
- une analyse des échanges opérateur-système, conduite par E. Kolmayer³².
- une analyse linguistique du corpus de questions, réalisée par la société GSI-ERLI³³.

Différents travaux en cours, notamment à l'ENSSIB, prolongent l'une ou l'autre de ces directions et approfondissent les résultats de l'enquête:

- étude faisant appel à la psychologie cognitive, sur l'organisation des connaissances des utilisateurs d'un OPAC (par E. Kolmayer).
 - analyse, dans le cadre d'un mémoire de DCB, du degré de correspondance pouvant exister entre les termes entrés par un utilisateur et le système d'indexation utilisé à la Médiathèque de la Cité des Sciences et de l'Industrie (par M. Encrenaz).
 - mémoire de DEA en Sciences de l'Information et de la Communication visant à constituer une typologie des usagers.
- etc.

Le présent mémoire s'inscrit dans la troisième direction de recherche, qui centre son étude sur des questions d'ordre linguistique et se propose d'étudier le dialogue qui s'instaure, parfois difficilement, avec l'OPAC.

31- Résultats de l'étude statistique. In [ECOLE 93], p. 45-74.

32- KOLMAYER, E., Analyse des échanges opérateur-système lors de l'interrogation de GEAC 9000. In [ECOLE 93], p. 75-85.

33- GSI-ERLI, Analyse linguistique du corpus de questions. In [ECOLE 93], p. 86-110.

3- Composante linguistique du projet

3-1 Objectifs

Sur le plan plus spécifiquement linguistique, le projet visait deux objectifs majeurs :

- étudier les caractéristiques linguistiques du dialogue qui s'établit entre l'utilisateur et le système
- caractériser les types de problèmes qui peuvent venir perturber ce dialogue homme/machine.

Il fallait, pour cela, disposer de données exploitables, ce qui n'était pas le cas jusqu'alors. (Les seules données recueillies étaient les listes de mots matières dont la recherche n'avait pas abouti. Ces listes, provenant de l'OPAC de la Bibliothèque Municipale de Valence et datant de 1988, ont été étudiées par P. Le Loarer³⁴).

3-2 Première analyse des données

L'analyse linguistique conduite par GSI-ERLI a porté sur deux questions du pré- et du postquestionnaire d'enquête, qui présentent toutes deux la caractéristique d'être des questions ouvertes :

- la question 24 du préquestionnaire (que nous appellerons PR24 dans la suite de ce travail) :

Décrivez assez précisément ce que vous voulez rechercher. Par exemple

Information sur la technologie des disques optiques

Dernier ouvrage d'ASIMOV traduit en français

- la question 11 du postquestionnaire (que nous nommerons par la suite P011) :

Comment auriez-vous souhaité formuler votre question dans un système idéal?

L'utilisateur a pu répondre à ces questions en langage naturel.

L'étude n'a pas porté sur l'ensemble du corpus, mais seulement sur les requêtes sujet ou sur l'aspect sujet de certaines requêtes pouvant également comporter une recherche de titre, d'auteur... Les réponses P011 identiques à PR24 ont été éliminées, de même que les réponses - nombreuses, nous le verrons - qui présentent un commentaire personnel de l'utilisateur ou un énoncé à caractère général.

Cette analyse s'est d'abord attachée à mettre en évidence les problèmes typographiques et orthographiques : GSI-ERLI a employé ses outils linguistiques Aleth pour tenter de redresser certaines erreurs. Une étude morpho-lexicale a ensuite conduit à proposer une classification des catégories morpho-syntaxiques repérées dans le corpus. Une analyse syntaxique rapide a enfin permis de dresser un *catalogue des principaux phénomènes syntaxiques relevés*. Selon Pierre Le Loarer, Directeur de projet pour la société GSI-ERLI, cette première approche a fait apparaître un *condensé de tous les problèmes de la linguistique automatique*.

34- LE LOARER, P., Opacité et transparence des catalogues informatisés pour l'utilisateur. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 69 et suiv.

3-3 Prolongements à l'analyse linguistique

L'étape suivante consiste à s'interroger sur la nature de la reformulation que l'utilisateur opère lorsqu'on lui laisse la possibilité de formuler sa question en langage naturel. Sylvie Lainé-Cruzel, qui s'est plus particulièrement intéressée à cette formulation libre de l'utilisateur, émet l'hypothèse que la formulation du postquestionnaire, lorsqu'elle existe, serait très proche de celle du préquestionnaire, voire identique : nous nous proposons, dans un premier temps, de compléter l'étude de la question P011, afin de confirmer ou d'infirmer cette hypothèse.

Nous constaterons cependant que l'étude de cette question ouverte ne suffit pas à rendre compte du dialogue utilisateur/système. C'est pourquoi nous confronterons, dans un deuxième temps, les diverses formulations du sujet que l'utilisateur a pu proposer au cours de l'enquête :

- formulation libre du préquestionnaire (PR24)
- formulation "en situation", c'est-à-dire termes entrés pendant la recherche
- formulation "idéale" du postquestionnaire.

IV- Modélisation d'une question en langage naturel

Deux types d'études, complémentaires, nous paraissent devoir être menées sur une question ouverte telle que la question 11 du postquestionnaire :

- une analyse syntaxique et sémantique des réponses P011, prolongeant celle conduite par la société GSI-ERLI, et destinée à faire apparaître des "types linguistiques" propres à une formulation exprimée en langue naturelle (ou quasi naturelle). Nous retiendrons pour cette analyse les axes suivants :

- * Types de réponses
- * Types de recherches
- * Domaines sémantiques
- * Etude morpho-lexicale
- * Analyse syntaxique

- une comparaison entre la question P011 et la formulation, également libre, exprimée dans la question 24 du préquestionnaire, afin de déterminer les évolutions ou les conservations de l'énonciation.

1- Typologie linguistique de la question P011

De nombreux utilisateurs - et nous reviendrons sur cette lacune de l'enquête électronique - ont omis de répondre à la question ; certains ont profité de sa formulation ouverte pour émettre des commentaires personnels sur le système et la Médiathèque, parfois fort intéressants mais peu utiles sur un plan purement linguistique. Françoise Belet notait déjà, à propos de l'enquête précédemment conduite à la Médiathèque, cette tendance marquée des utilisateurs à sortir du cadre des questions et à émettre ce qu'elle appelle des *commentaires sauvages*³⁵. La présence de questions ouvertes dans cette enquête ne pouvait manquer d'engendrer les mêmes comportements, dans des proportions importantes puisque 148 réponses seulement sur 598 se révèlent exploitables.

Dans la suite de ce travail, chaque enregistrement sera désigné par un nom ainsi composé :

- 2 ou 3 lettres désignant le poste sur lequel a été enregistré le questionnaire ou la session
- un nombre à 2 chiffres correspondant au jour de l'enregistrement (toutes les sessions ont été enregistrées en Juin 92).
- un signe de ponctuation : "."
- un chiffre ou un nombre à 2 chiffres, représentant le numéro d'ordre de la session sur ce poste et pour cette journée.

Ainsi la session TTA04.5 est la cinquième session enregistrée le 4 Juin sur le poste TTA. Nous avons repris, par commodité, les noms des fichiers de synthèse correspondant à chaque session, fichiers d'analyse automatique conçus par Mohamed Hassoun³⁶.

35- BELET, F., Id., p. 27.

36- HASSOUN, M., Traitement informatique des résultats de l'enquête. In [ECOLE 93], p. 113-135.

Nous avons distingué au moyen de caractères en script les termes et phrases entrés par les utilisateurs dans les questionnaires, et plus tard dans les sessions. Ils sont reproduits intégralement et sans correction de notre part : la typographie (majuscules/minuscules) et surtout l'orthographe adoptées par l'utilisateur ont été conservées.

On remarquera que les utilisateurs ont entré leur texte - et leurs termes de recherche - en adoptant une typographie tantôt riche, tantôt pauvre : en majuscules, en minuscules ou en typographie mixte ; avec accent ou sans accent... Nous n'avons pas pris ce phénomène en compte dans nos analyses, pour diverses raisons : l'adoption de telle ou telle typographie ne semble pas toujours correspondre à un choix linguistique de l'utilisateur (qui sait ou ignore, par exemple, qu'un nom propre commence par une majuscule), mais plutôt à son niveau de maîtrise du clavier. D'autre part, des problèmes techniques, liés aux PC utilisés puis au transcodage des données, ont pu modifier la graphie de départ, problèmes dont nous ne pouvons évaluer l'ampleur. GSI-ERLI analyse ces problèmes dans ses propositions de redressement³⁷.

1-1 Types de réponses

Le tableau suivant offre une répartition des 598 réponses ou non-réponses en 5 grands types qu'il convient de présenter :

- Le terme "recherches spécifiques" désigne toute réponse contenant une ou plusieurs recherches , quel qu'en soit le mode (auteur, titre, sujet, etc.).

AC11.6 - *ou puis je trouver des livres dans la collection découvertes gallimard notamment sur les découvertes polaires*

- Les "recherches générales" réunissent des formulations telles que :

AC11.11 - *comment puis je trouver ce livre et dans quel rayon*

TTA09.3 - *avez vous telle revue?*

- Les "justifications de démarches" ne comportent pas de mention d'un thème de recherche, mais la manière dont on pourrait conduire ou améliorer le processus de recherche. (Nous avons inclus dans ce type les suggestions très fantaisistes de certains utilisateurs, qui ont respecté la consigne et imaginé un système vraiment "idéal").

UG13.4 - *DU GENERAL AU PLUS DETAILLE CITER LE DOMAINE DE RECHERCHE, LE SUJET, AVEC DES MOTS CLES*

ME09.13 - *systeme vocal*

- Les remarques personnelles regroupent à la fois les appréciations et critiques des usagers :

ME27.8 - *IL N'Y A PAS DE SYSTEME IDEAL*

et les réponses du type "idem" ou "de la même façon".

- Le cinquième groupe se compose des vides et des réponses du type "?", "pas de réponse" ou "je ne sais pas".

Quelques sessions ambiguës - 13 au total - n'ont pu être classées dans l'un ou l'autre de ces types, mais il ne s'agit jamais de recherches spécifiques.

37- GSI-ERLI, Analyse linguistique du corpus de questions. In [ECOLE 93], p. 90-92.

Synthèse des types de réponses par poste de consultation :

	rech. spéc.	rech. gén.	just. dém.	rqs. perso.	pas de rép.	autre
ACC (84)	26	5	14	12	26	1
ME (212)	43	7	36	51	70	5
UG (187)	42	0	35	34	71	5
TTA (110)	34	1	17	22	34	2
MHS (5)	3	0	1	0	1	0
Tot. (598)	148	13	103	119	202	13
%	24,7	2,2	17,2	19,9	33,8	2,2

La suite de notre étude des réponses P011 ne prendra en compte que les recherches spécifiques, les seules qui proposent une énonciation du thème de recherche intéressant l'utilisateur. Il faut cependant d'ores et déjà remarquer le taux particulièrement élevé de non-réponses, d'origines diverses : l'utilisateur satisfait de sa recherche ne voit pas l'utilité de reformuler son sujet ; certains ne comprennent pas la question ; d'autres encore ne parviennent pas à imaginer *un système idéal*.

1-2 Types de recherches

La plupart des 148 recherches "spécifiques" présentent une recherche:

- par auteur (Aut)
- par titre (Tit)
- par sujet (Suj)

ou associent 2 modes de recherche, voire davantage : multicritères (Mult). L'une fait appel à une catégorie : "science-fiction".

Une réponse a été décomposée en deux parties, car elle contient deux recherches différentes, l'une par titre et l'autre par sujet et éditeur : ME14.4. D'où un total de 149 sujets.

Quelques réponses présentent des ambiguïtés (Amb) et n'ont pas été classées. Ainsi la session UG04.12 : *préAO titre ou sujet...*

Ce classement reflète l'écrasante majorité des recherches par sujet, que des enquêtes antérieures à celle de PARINFO-OPACs avaient déjà mise en évidence. Une lecture des réponses que nous avons classées en "recherche générale" ou "justification de démarche" le confirme : 40 utilisateurs sur 62 choisissent la recherche par sujet.

Tableau récapitulatif des types de recherches par poste :

	Aut	Tit	Suj	Cat	Mult	Amb.	Tot.
ACC	2	1	16	1	6	0	26
ME	1	1	38	0	3	1	44
UG	1	1	29	0	9	2	42
TTA	0	0	29	0	5	0	34
MHS	0	0	3	0	0	0	3
Tot.	4	3	115	1	23	3	149
%	2,7	2	77,2	0,7	15,4	2	100

1-3 Typologie sémantique

L'enquête de GSI-ERLI, sans proposer de typologie, concluait à une *nette prédominance des sujets informatiques dans le corpus étudié*, ce que l'on pouvait d'ailleurs prévoir, étant données la spécificité du fonds de la Médiathèque et son orientation vers les nouvelles technologies. Elle estimait qu'une typologie poussée serait intéressante si elle était conduite à un niveau relativement fin.

Il nous a paru cependant utile de montrer non pas l'importance prévisible des domaines scientifiques et techniques, mais le nombre, plus élevé que prévu, des recherches inadaptées au fonds de la Médiathèque. Ces formulations de requête font le plus souvent suite à des recherches correctement conduites, mais décevantes. Certains utilisateurs conçoivent difficilement qu'un auteur particulièrement connu ou qu'un thème d'actualité soit absent du fonds.

Cette remarque en témoigne : (réponse P011 - session TTA17.12)
LA MEDIATHEQUE N'ETANT PAS SPECIALISEE EN SCIEN.sociales et humaines je comprends bien que des ouvrages specialisees ne s'y trouvent pas. Mais "l'éducation familiale" doit etre promue..

Notre classement comprend la totalité des 148 recherches dites "spécifiques", pas uniquement les 115 réponses proposant un sujet.

37% des recherches ne paraissent pas adaptées au fonds. En réalité, ce chiffre est sans doute trop élevé : certains sujets classés en sciences économiques et sociales portent par exemple sur l'entreprise, domaine dans lequel la Médiathèque a constitué un fonds de base. Le travail actuellement réalisé par M. Encrenaz sur l'adéquation entre les termes entrés par l'utilisateur et le référentiel de la Médiathèque et sur les causes d'échec de certaines recherches devrait permettre d'évaluer l'importance réelle des recherches inadaptées.

Tableau récapitulatif :

L = littérature et philosophie
 A = arts, sports
 H = histoire, géographie
 So = sciences économiques et sociales
 Sc = sciences, médecine
 T = techniques non informatiques
 I = informatique
 NC = non classée

	L	A	H	So	Sc	T	I	NC
Acc	3	1	4	5	4	4	3	2
MEIJ	1	3	0	2	13	8	15	1
UGO	3	2	3	5	17	2	7	3
TTA	1	1	1	20	3	3	4	1
MHS	0	0	0	0	3	0	0	0
Tot.	8	7	8	32	40	17	29	7

1-4 Analyse morpho-lexicale

Nous reprendrons les catégories dégagées par GSI-ERLI, mais en tentant de chiffrer l'importance des phénomènes repérés.

1-4-1 Nombre de mots par réponse

Chaque réponse comporte en moyenne 6,5 mots : les réponses à la question P011 semblent plus courtes que celles concernant la question PR24, alors que l'on pouvait espérer le contraire.

La majorité des unités lexicales consiste en noms communs et adjectifs. On note cependant la présence, moindre par rapport à PR24, de noms composés, de noms propres, de sigles, d'abréviations et de numériques, ainsi qu'une utilisation supérieure des verbes.

La totalité des sujets est exprimée en français, à une exception près: ME27.4 - *Mathematical models*.

1-4-2 Catégories morpho-lexicales

Les **verbes** constituent une catégorie importante. On en relève 67, le plus souvent conjugués :

- 42 au mode indicatif
- 13 au mode infinitif
- 6 au mode participe
- 5 aux modes subjonctif ou conditionnel
- 1 au mode impératif

Ces verbes jouent un rôle introducteur dans la formulation du thème de recherche :

AC12.2 - *je cherche des documents concernant e.KANT*

UG17.18 - *ou peu on trouver des documents sur le systeme solaire*

ME23.10 - *Trouve moi tous les documents sur les contrats de partenariat*

Ils constituent donc des termes "vides" par rapport à une éventuelle reformulation du sujet.

On relève un cas unique de **mot composé** : *science-fiction*, dans la session AC11.7.

Un quart des sessions comporte des **noms propres** (42) :

- noms de personnes, d'auteurs : 15
- noms d'éditeurs : 3
- titres : 2
- lieux géographiques : 6
- systèmes ou logiciels informatiques : 13
- autres : 3 (jeu, personne morale, nom historique).

Certaines réponses utilisent des **sigles et acronymes**, parmi lesquels *PME, IGBT, RMI, SNCF...*(13)

On note 9 cas d'**abréviations**, dont 3 semblent dûs au système et non à l'utilisateur :

- abréviation d'un prénom : *e.KANT ; C RAULT*
- abréviations de noms communs : *math (proba) exo ; maths ; pb* pour publics; (*asse* pour assembleur, *etrange* pour étrangers et *bancair* pour bancaires)
- amalgame entre sigle et abréviation : *PréAO* (UG04.12).

4 réponses incluent des **numéraux** ou des **cardinaux** (*windows 3.1 ; 1982; 1800 et 1920 ; 1ere*), 5 comprennent des **signes mathématiques** : +, =, (), &.

Etudier l'ensemble des noms communs et adjectifs utilisés n'apporterait pas nécessairement de données nouvelles sur le corpus. Il nous semble cependant intéressant de signaler l'importance d'un champ sémantique spécifique, celui des noms de support, qui fait son apparition dans les formulations en langage libre.

La plupart de ces termes (67 au total) sont des noms communs, au singulier ou au pluriel, vides de sens par rapport au fonds de la Médiathèque. Il est en effet inutile de mentionner que l'on cherche un livre ou un document.

En revanche, des termes tels que "revue" ou "manuel" apportent une information supplémentaire sur la nature du support recherché : marqués sur le plan sémantique, ils devraient être pris en compte si la réponse était traitée par un analyseur. Mais ces termes - que nous avons signalés en italiques - ne sont pas les plus nombreux.

Faut-il d'autre part distinguer entre les demandes d'information(s) et les demandes de livres ou de documents ? L'utilisateur sait qu'il ne se trouve pas devant un système capable de lui fournir de l'information primaire, mais seulement une information secondaire.

- livre / livres : 18
- document / documents : 15
- ouvrage / ouvrages : 9
- références : 2
- information / informations : 5
- autres :
 - . documentations : 1
 - . bouquins : 1
 - . revue / revues : 5
 - . roman : 1
 - . manuel d'utilisation : 1
 - . répertoire : 1
 - . dictionnaire : 1
 - . guide / guides : 2
 - . atlas : 1
 - . texte de loi : 1
 - . tableau : 1
 - . photographies : 1
 - . diapositives : 1

Ces notations de support constituent une tentative d'explicitation du sujet de recherche en langage naturel, mais une tentative pas nécessairement spontanée ; la formulation même de la question PR24 a très probablement influencé l'utilisateur :

"Décrivez assez précisément ce que vous voulez rechercher. Par exemple :
Information sur la technologie des disques optiques
Dernier ouvrage d'ASIMOV traduit en français"

1-5 Analyse syntaxique

1-5-1 Catégories syntaxiques

La catégorie majoritairement représentée, et nous rejoignons en cela l'analyse de GSI-ERLI, est celle des **syntagmes nominaux complexes** : 41,5 % des réponses, soit 62 cas sur 149. (Rappelons que la session ME14.4 comprend 2 formulations de sujets).

ME06.10 - *informations sur les series de fonction*
 UG17.32 - *matabolisme des graisses dans le foie*
 TTA24.4 - *cryptage de donnees en matiere de cartes bancair*

Celle des **syntagmes nominaux simples** - du type uniterme, Nom + Adjectif, Nom commun + Nom propre - et des syntagmes simples coordonnés comprend 22 occurrences, soit 15 % du corpus.

ME09.15 - *autocad*
 UG05.4 - *semiologie chirurgucale*
 TTA26.3 - *logiciel reflex*

Nous n'avons identifié que 2 **phrases complexes** (AC11.15 et AC11.18), chiffre qui semble inférieur à celui de PR24.
 AC11.15 - *savoir comment ca se passe une intervention chirugicale de l otite chronique*

Un autre groupe important est constitué de **phrases simples**, de différentes structures :

- structures interrogatives : 27, soit 18 %
UG12.8 - *existe-t-il des documents sur les musées étrange*
- structures déclaratives : 17, soit 11,5 %
TTA06.12 - *le sujet est climat social dans l'entreprise*

2 structures interrogatives ont été exclues de ces comptages car elles constituent des **énoncés agrammaticaux** :

- ME19.2 - *Quels sont les documents relatifs aux mémoires informatiques possédez-vous?*
- ME23.4 - *quels rayons se situant à quels étages peut-on trouver des documents de maths*

11,5 % des réponses (17 cas) s'expriment sous la forme de **structures énumératives**, c'est-à-dire de juxtapositions de termes, elliptiques des articles, des prépositions et des coordinations :

- ME19.5 - *THOM CATASTROPHES*
- UG10.6 - *LIVRE-anglais-photographie computerise*

1-5-2 Quelques phénomènes syntaxiques

La formulation de la question en langue naturelle oblige à traiter un certain nombre de phénomènes et d'irrégularités linguistiques que les langages d'interrogation contrôlés visent précisément à éliminer. Nous ne reviendrons pas sur les nombreuses fautes de frappe et d'orthographe, pour lesquelles GSI-ERLI a proposé une typologie et des possibilités de redressement³⁸. Les deux principales difficultés que nous avons pu repérer dans le corpus des réponses proviennent d'une part des tournures elliptiques, d'autre part des structures de coordination.

Les **ellipses** sont particulièrement nombreuses dans les structures énumératives, mais on en repère également dans certaines phrases simples (25 cas) :

- ellipse des prépositions et/ou des articles

- AC17.19 - *répertoire associations protection nature*
- ME13.1 - *recherche sur système OCCAM*
- TTA10.3 - *organigramme sncf*

- formulation de la requête sur le modèle d'une petite annonce, en élidant sujet, prépositions et articles.

- AC16.5 - *recherche tout document sur lutherie*
- UG18.13 - *recherche information sur le turbo pascal*
- UG26.4 - *cherche documents gastronomie*

- ellipse de la tournure interrogative (1 cas) :
ME21.5 - *livres de l'auteur goldstein disponibles?*

38- GSI-ERLI, Analyse linguistique du corpus de questions. In [ECOLE 93], p. 94-102.

14 réponses se présentent sous la forme d'une **structure coordonnée** :

- coordination de deux unitermes

AC19.3 - *eglises et paroisses*

- coordination d'un uniterme et d'un syntagme nominal

AC27.9 - *energie et pays en voie de developpement*

UG23.10 - *sujet veterinaire et revue*

- coordination de deux syntagmes nominaux

UG13.3 - *dernier ouvrage d'ALBERT JACQUARD et autres ouvrages sur ce thème*

- coordination de deux syntagmes prépositionnels

TTA07.2 - *recherche sur l'enfant et l'urbanisme*

La coordination est le plus souvent exercée par "et", mais on note aussi une utilisation de "ou", ainsi que des signes "+" et "&" :

TTA11.5 - *recyclage puis aluminium ou metaux*

ME05.21 - *DOMAINE unix + auteur*

UG16.10 - *BASE & donnees*

L'emploi de la coordination génère parfois des ambiguïtés syntaxiques, comme dans les exemples suivants :

AC25.3 - *sponsoring et mecenat pour les PME*

TTA20.10 - *information sur la concurrence industrielle et sur la reglementation de la concurrence en europe (s'agit-il de "concurrence industrielle" ou de "concurrence industrielle en europe" ?)*

Il semble cependant que le cas soit moins fréquent en P011 qu'en PR24.

La coordination engendre aussi des ambiguïtés sur le plan sémantique, ambiguïtés liées aux multiples interprétations possibles du connecteur "et".

La confrontation de ces trois énoncés suffit à le prouver :

AC19.3 - *eglises et paroisses*

AC16.12 - *communications et unix*

UG27.1 - *bibliographie et sciences*

Dans le premier cas, l'utilisateur recherche des documents portant sur les églises et/ou sur les paroisses ; dans le second, il souhaite vraisemblablement croiser les deux termes : le connecteur s'apparente ici au "et" de la recherche booléenne. Le fait est encore plus apparent dans le troisième exemple (PR24 confirme que l'utilisateur recherche des documents de bibliographie scientifique).

Outre l'ellipse et la coordination, on repère quelques structures qui pourraient créer des difficultés si on souhaitait traiter ces formulations de manière automatique :

- Anaphores (3) :

AC13.18 - *excusez moi de vous deranger, mais vous savez ou sont les bouquins de zola, marcel de son prenom*

UG13.3 - *dernier ouvrage d'ALBERT JACQUARD et autres ouvrages sur ce thème*

UG09.10 - *je cherche les techniques de construction des serres et les principes*

- Structures apposées, qui visent à reformuler le sujet (5)

AC11.6 - *ou puis je trouver des livres dans la collection découvertes gallimard **notamment sur les découvertes polaires***

Certaines de ces précisions figurent entre parenthèses :

UG14.3 - *recherche d'un guide géographique (region paris.)*

- Inversions de termes (3) :

ME11.7 - *math 1ere (proba) exo*

ME13.11 - *electronique interfaces parallele technique*

ME27.4 - *Mathematical models*

- des constructions syntaxiques familières ou incorrectes, parfois dues à un emploi erroné de certaines prépositions :

AC11.7 - *pourriez vous me renseigner sur un roman de science-fiction!*

AC12.4 - *quelles sont les documents qui peuvent traiter sur la gestion des ressources humaines dans les PME?*

AC11.15 - *savoir comment ca se passe une intervention chirurgicale de l otite chronique*

Ce corpus de formulations en langage naturel offre en définitive deux tendances opposées : l'une consiste à énumérer des termes sans préciser le lien qui les unit, un système idéal devant se charger de les apparier ; d'autre part, la reformulation est l'occasion de lier ces termes dans des structures de préférence simples, qui n'apportent rien sur le plan sémantique : on utilise à cet effet des verbes introducteurs, voire des formules de politesse (dont certaines peuvent cependant s'interpréter comme s'adressant à l'enquêteur qui, par-delà le système, relira ou traitera le questionnaire électronique).

Entre ces deux tendances, 56,5 % des utilisateurs s'en tiennent à des syntagmes, soit conformes à ceux qu'ils sont susceptibles d'entrer durant une session, soit élargis, mais le plus souvent au moyen de termes indiquant la nature du support recherché. Cette étude des structures syntaxiques semble donc bien confirmer que la majorité des utilisateurs n'a pas reformulé son sujet à l'issue de sa session, mais s'est laissée influencer par ses formulations précédentes. D'où la nécessité de confronter ces énoncés avec les réponses issues du préquestionnaire.

2- Comparaison PR24 / P011

Cette étude confronte les 148 recherches spécifiques identifiées parmi les réponses P011 aux réponses PR24 correspondantes.

Elle nécessite d'éliminer d'emblée 9 sessions:

- 2 ne possèdent pas de réponse à la question PR24 (UG17.32 et TTA04.4).

- 7 ne présentent aucun lien, ni syntaxique, ni sémantique : l'utilisateur a modifié son sujet de recherche d'un questionnaire à l'autre. (AC13.8 ; AC23.4 ; ME24.2 ; UG07.7 ; UG09.15 ; UG21.1 ; TTA07.7)

Exemple : TTA07.7

PR24 = *ELECTRONIC PRATIQUE*

P011 = *QU AVAIS VOUS SUR LE SPORT?*

53 enregistrements font état d'un lien syntaxique fort - identité ou inclusion - entre les réponses formulées aux deux questionnaires électroniques, 70 d'un lien syntaxique marqué par la reprise d'un ou plusieurs termes et 16 d'un lien sémantique mais non syntaxique.

Notre objectif est de tenter d'identifier les cas de "reformulation", au sens que P.W. Dalrymple donne à ce mot :

*Reformulation is conceptually defined as an interactive refinement of the term or terms used to locate bibliographic items in a catalog*³⁹.
(La reformulation se définit conceptuellement comme un processus interactif d'affinage du ou des termes utilisés pour retrouver des références bibliographiques dans un catalogue.)

2-1 Identité ou inclusion

12 réponses à la question P011 s'avèrent strictement ou pratiquement identiques à celles de la question PR24 (à un article, une faute de frappe ou des guillemets près).

TTA16.2 - PR24 = P011 = COMPOSANTS DIVERS DU SANG
UG05.4 - PR24 = *semiologie chirurgicale*
P011 = *semiologie chirurgicale*

18 formulations sont extraites des réponses à la question PR24. Dans la plupart des cas (13), l'énoncé reste stable sur le plan sémantique : l'utilisateur a abandonné des mots "vides".

UG23.4
PR24 = *LIVRE CONCERNANT LA PHYSIOLOGIE DES OISEAUX*
P011 = *physiologie des oiseaux*

Mais elle peut aussi s'accompagner d'une extension ou d'une réduction du sujet. On peut supposer qu'elle prend alors en compte les difficultés rencontrées au cours de la session.

TTA11.8
PR24 = *DOSSIER COMPLET SUR LA FRANCHISE EN EUROPE*
P011 = *FRANCHISE*

Dans la session UG13.14, la perte du mot "societes" rend le sujet incompréhensible :

PR24 = *comptabilite des societes fusion*
P011 = *comptabilite fusion*

On relève 23 occurrences dans lesquelles c'est la réponse PR24 qui est incluse dans P011. Comme précédemment, la règle est la stabilité sémantique (16 cas).

AC11.17
PR24 = *programmation windows 3.1*
P011 = *des livres de programmation sue windows "3.1*

39- DALRYMPLE, P.W. Retrieval by reformulation in two library catalogs : toward a cognitive model of searching behavior. *Journal of the American Society for Information Science*, 1990, Vol. 41, n°4, p. 274.

2 réponses induisent cependant une spécification du sujet (AC25.3 et UG14.3), tandis que 4 apportent des précisions sur l'auteur ou le type de document recherché :

AC11.18

PR24 = *structure des systemes dynamiques*

PO11 = *je cherche STRUCTURE DES SYSTEMES DYNAMIQUES de JM SOURIAU, pourriez vous m aider?*

L'une des sessions, AC11.15, présente un développement du sujet, que l'on peut interpréter comme une reformulation :

PR24 = *otite chronique*

PO11 = *savoir coment ca se passe une intervention chirurgicale de l otite chronique*

A cette exception près, la formulation reste très proche d'un questionnaire à l'autre.

2-2 Reprise de termes (70 cas)

On pourrait supposer que l'utilisateur, en sélectionnant des termes déjà énoncés et en les recomposant, a cherché à reformuler son sujet. En réalité, un certain nombre de sessions soit présentent une certaine stabilité de l'énonciation (24 cas), soit focalisent sur un thème de recherche alors que PR24 en comportait plusieurs (10), soit encore intègrent une précision sur le domaine ou le type de document recherché (7).

Exemple de stabilité : AC14.5

PR24 = *Technologies de transfert et stockage d'image avec identification sur des critères non descriptifs*

PO11 = *Quels sont les documents publiés sur le transfert et le stockage de l'image, ainsi que la caractérisation des images sur des critères autres que descriptifs*

Exemple de réduction de thème : ME12.15

PR24 = *TRI QUICKSORT EN LANGAGE C ET CALCUL DE COMPLEXITE DE PROGRAMMES*

PO11 = *algorithme de tri quicksort en langage c*

Exemple de précision du type de document souhaité : UG28.8

PR24 = *documents sur les ours bruns*

PO11 = *photographies sur les ours bruns*

Dans les 29 occurrences restantes, la reprise de certains termes de PR24 s'accompagne d'une extension (7 cas), d'une restriction (7 cas) ou d'un déplacement sémantique du sujet (15) ; cette dernière notion est parfois difficile à cerner. Seules ces 29 occurrences peuvent comporter une reformulation du sujet, sans que ce soit systématique.

Extension du sujet : UG26.4

PR24 = *gastronomie europeenne sciences des alimentqs*

PO11 = *cherche documents gastronomie*

Restriction : ME14.12

PR24 = *Information sur les jeux de strategie (RISIKO)*

PO11 = *Liste des titre des livres sur le jeu du Risiko*

Déplacement : TTA07.2
 PR24 = *URBANISME ET ENVIRONNEMENT*
 PO11 = *recherche sur l'enfant et l'urbanisme*

La confrontation des réponses PR24/PO11 permet d'établir, pour ces 29 occurrences, que l'utilisateur a réorienté son sujet, à la fois sur le plan syntaxique et sémantique. Mais elle ne permet pas d'apprécier les causes de ce phénomène : doit-on y voir une influence de la session, au cours de laquelle l'utilisateur aurait compris que son sujet était trop précis, trop vague ou inadapté ? Son choix reflète-t-il alors les propositions que le système a émises durant la session ? Autrement dit, y a-t-il ou non apprentissage linguistique durant une session ? Une simple comparaison entre les deux questionnaires électroniques ne permet pas de répondre à ces questions.

2-3 Lien sémantique (16 cas)

Les enregistrements classés dans cette catégorie se caractérisent par le fait que PO11 et PR24 ne possèdent aucun terme commun, mais portent sur un même domaine sémantique.

TTA24.6

PR24 = *evolution des sciences humaines dans le batiment*

PO11 = *modèles d'organisation du travail en travaux pb*

Dans ce cas précis, l'utilisateur a véritablement reformulé son sujet, de façon plus explicite qu'en PR24.

Dans la plupart des cas cependant, l'utilisateur a modifié le sujet de départ :

- soit en le spécifiant

ME09.11

PR24 = *geometrie dans l'espace, polyedres*

PO11 = *informations sur dodecaedre*

- soit en le généralisant

UG09.8

PR24 = *l'homme face au risque technique l'harmattan &çç*

PO11 = *envbironnement*

- soit en le déplaçant

ME11.7

PR24 = *formulations, définitions d'optimpisation et de convexite*

PO11 = *math 1ere (proba) exo*

Si nous pensons que les cas de spécifications et d'extensions du sujet peuvent constituer des tentatives de reformulation, il nous semble difficile d'étendre cette conclusion aux cas de déplacements du sujet : la reformulation, si elle s'accompagne par définition de modifications lexicales et syntaxiques, doit conserver la "position sémantique" du sujet par rapport à l'énoncé de départ : ce n'est pas le cas dans les sessions ME11.7, UG20.10, TTA20.10.

La comparaison des sujets formulés en PR24 et en P011 peut donc se résumer comme suit :

- 2 enregistrements sans réponse PR24
- 7 réponses P011 sans rapport avec PR24
- 12 cas d'identité entre les 2 formulations
- 41 cas d'inclusion, parmi lesquels 7 spécifications ou extensions du sujet
- 24 reprises de termes, avec stabilité de l'énonciation
- 10 reprises, avec réduction du nombre de thèmes de PR24 à P011
- 7 reprises, avec précision du domaine ou du type de document
- 29 reprises, avec restrictions, extensions ou déplacements du sujet
- 9 enregistrements sans termes communs, mais présentant une stabilité, une réduction ou une extension du domaine sémantique de départ
- 7 enregistrements sans termes communs, présentant un déplacement sémantique du sujet.

Au terme de cette première étude, nous serions tentée de conclure que 45 enregistrements sur 148, soit 30 %, contiennent une tentative de reformulation du sujet. Ce nombre ne représente que 7,5 % des 598 sessions enregistrées, pourcentage très faible mais révélateur à la fois des limites de l'enquête électronique et des difficultés de l'utilisateur.

2 réponses constituent de véritables reformulations, au sens que nous donnons à ce terme : AC11.15 et TTA24.6. Dans ces cas de figure, l'utilisateur a conservé l'extension sémantique de son sujet de départ, mais en a proposé une nouvelle expression linguistique. Dans la plupart des autres cas, la reformulation s'accompagne d'une spécification, d'une extension, voire d'un déplacement sémantique du sujet : la frontière entre reformulation et nouveau sujet devient alors difficile à établir. Notre classification des enregistrements peut parfois sembler subjective...

L'étude précédente comporte une autre limite : elle ne permet pas de comprendre les motivations de l'utilisateur. Pourquoi a-t-il éprouvé le besoin d'étendre ou de réduire son domaine de recherche ? La conservation totale ou partielle d'une formulation de départ - 94 enregistrements sur 148, 63,5 % des cas - signale-t-elle que l'utilisateur est satisfait de son énoncé et de la session conduite à partir de cet énoncé ou traduit-elle au contraire son incapacité à reformuler le sujet de départ ?

Pour tenter de répondre à ces interrogations, il est nécessaire de s'intéresser au déroulement même de la session, chaînon manquant entre les deux questionnaires électroniques.

CHAPITRE 2

Formulation de la requête,
du préquestionnaire au postquestionnaire

Notre objectif consiste ici à étudier l'ensemble des formes linguistiques que peut revêtir une question au cours d'une session donnée : proposition initiale de l'utilisateur, telle qu'elle est fournie dans la question 24 du préquestionnaire et dans la première étape de la session, mais aussi évolution de cette formulation au cours de la session, et formulation "dans un système idéal", lorsqu'elle existe.

Dans quelle mesure la relation avec le système modifie-t-elle la formulation proposée par l'utilisateur et le sujet même de sa recherche ? Peut-on mettre en évidence des stratégies d'interrogation, et en particulier des stratégies de reformulation ? Pour tenter de répondre à ces questions, il est nécessaire d'étudier l'ensemble d'une session, du préquestionnaire au postquestionnaire.

I- Méthodologie

Nous n'aurions pu conduire ce type d'étude sur l'ensemble des sessions enregistrées à la Médiathèque de La Villette. C'est pourquoi nous avons constitué un corpus, en choisissant pour critère le nombre de références effectivement visualisées par l'utilisateur au cours de son interrogation. Nous avons donc retenu les 154 sessions durant lesquelles l'utilisateur a affiché au moins 3 notices. L'étude portera sur un quart environ des 598 sessions enregistrées en Juin 92.

1- Constitution d'un corpus

1-1 Critère de sélection

Le choix de ce critère répond à deux objectifs spécifiques:

Les sessions étudiées, sans nécessairement répondre aux attentes de l'utilisateur, lui ont fourni des références suffisamment intéressantes sur le plan linguistique pour être affichées. Ce critère objectif permet ainsi de dépasser l'appréciation de l'utilisateur, notamment le jugement qu'il porte sur sa propre interrogation dans la question 6 du postquestionnaire : *Cette recherche vous a paru très facile, facile, difficile, impossible*. En effet, nous avons pu constater que les réponses à cette question étaient très subjectives, variant en fonction du niveau de formation de l'utilisateur aux OPACs, de sa connaissance du domaine, du degré de précision et de spécialisation de sa question...

Les sessions retenues devraient aussi permettre de déterminer si le matériau linguistique contenu dans les références affichées a ensuite été réutilisé - en quelque sorte "réinjecté" - dans une éventuelle reformulation de la question, au cours d'une étape suivante de la recherche. La visualisation de références permet-elle une reformulation linguistique, en est-elle une étape nécessaire ou facilitante ?

1-2 Sessions non retenues

Nous n'avons pas eu la possibilité matérielle, étant donné le nombre de données et le temps qu'une telle étude aurait représenté, de nous intéresser aux 444 sessions que nous avons écartées.

Il faut cependant noter que ces enregistrements (3/4 du corpus total) ont donné lieu à un nombre d'affichages de références très faible, inférieur ou égal à 2 :

- 121 sessions sans affichage
- 202 à 1 affichage
- 121 à 2 affichages

On peut alors se demander si le corpus retenu ne concentre pas toutes les sessions linguistiquement "difficiles", les autres ayant ou bien totalement échoué, ou bien immédiatement satisfait l'utilisateur. Divers facteurs nous paraissent réfuter cette hypothèse, qui ne serait d'ailleurs pas contraire à nos objectifs d'étude :

La démarche de l'utilisateur ne dépend pas nécessairement du nombre d'affichages. Nombreuses sont les sessions parmi les 154 retenues où une même notice a été appelée plusieurs fois, sans interaction linguistique supplémentaire avec le système. Une démarche identique aurait pu être mise en oeuvre et ne donner lieu qu'à un affichage.

L'absence d'affichage ne permet pas de savoir à priori si la session contient ou non une interaction, et si les résultats éventuellement affichés sous forme abrégée (auteur/titre) ont satisfait l'utilisateur. (Celui-ci pouvait en effet consulter le catalogue pour vérifier qu'un ouvrage donné figurait bien dans le fonds de la Médiathèque. A l'exception d'ailleurs de quelques échecs flagrants, seul l'auteur d'une recherche pourrait exposer les motifs qui l'ont poussé à poursuivre ou à interrompre son interrogation.)

La présence d'un ou plusieurs affichages constitue, en revanche, une preuve qu'il y a eu au moins une interaction avec l'OPAC, c'est-à-dire une entrée de terme traitée par le système. Le critère des 3 affichages est ensuite arbitraire, mais il avait le mérite de fournir un corpus de taille raisonnable.

Nous n'avons pas cherché à sélectionner des recherches faciles ou difficiles à priori - bien que cette distinction présente un intérêt dans un autre contexte d'étude - mais des recherches que nous supposons capables de conduire l'utilisateur à modifier sa démarche linguistique et par suite son interaction avec le système. Notre sélection ne nous permet cependant pas d'émettre des hypothèses sur la qualité de l'échange linguistique réalisé dans les autres sessions. Cela constitue peut-être une de ses limites.

1-3 Organisation des données

Rappelons que le lecteur trouvera dans l'annexe 2 (volume 2) de ce mémoire quelques données relatives à chacune des sessions qui ont servi de support à l'étude linguistique, ainsi que leur répartition dans le plan de classement que nous adoptons dans ce chapitre :

- Sessions en une étape et un mode de recherche (46)
- Sessions en plusieurs étapes et un mode de recherche (39)
- Sessions en plusieurs étapes et deux modes de recherche (31)
- Sessions en plusieurs étapes et 3/4 modes de recherche (6)
- Sessions présentant des ambiguïtés (6)
- Sessions en plusieurs recherches indépendantes (26)

Nous l'invitons également à se reporter au tableau des codes et abréviations utilisés, en annexe 4.

2- Obstacles et limites

L'utilisation stricte du nombre de références affichées figurant dans le fichier de synthèse de chaque session aurait dû nous amener à retenir 4 sessions supplémentaires : ME26.7, UG17.19, TTA23.7 et TTA25.3 comportent a priori 3 références affichées, mais la première notice résulte en fait de la recherche précédente et a été comptabilisée par erreur avec ces sessions. Cela ramène le nombre de références affichées à 2.

Le nombre de notices affichées prend en compte la totalité des affichages, y-compris les multiples sélections d'une même notice. Dire qu'une session, par exemple AC11.7, a donné lieu à 4 affichages ne signifie pas que 4 notices différentes ont été visualisées : en réalité, 2 références ont été affichées 2 fois chacune. Chaque choix de référence ou de la commande ABR est comptabilisé.

Prenons un cas-limite : la session ME10.3. La recherche conduite à partir des sujets *préparation des hors d'oeuvre, réfrigération, préparation culinaire* et de l'auteur *Guyot, Michel* a effectivement donné lieu à 5 affichages (écrans 23,26,30,32,41), mais il s'agit de la même référence : *La prérefrigération : étude économique*.

Les statistiques sur les affichages de références ne tiennent pas compte des notices complètes, mais uniquement des notices abrégées.

Tout choix de référence à partir d'une liste de titres donne automatiquement lieu à un affichage sous forme abrégée, de même que l'utilisation de la commande ABR ; une notice abrégée est également affichée automatiquement si le système ne trouve qu'une réponse à la question de l'utilisateur. Une notice complète s'obtient en utilisant la commande COM.

Une notice abrégée comprend les informations suivantes :

- auteur
- titre
- mention d'édition
- éditeur
- support
- localisation cote
- disponibilité du document

La notice complète ne fournit pas la localisation, mais ajoute aux données bibliographiques de base une description physique du document, l'ISBN ou ISSN, un résumé, le sujet.

Il peut cependant arriver qu'une référence soit directement affichée sous forme complète : lorsque l'utilisateur feuillette des notices avec les commandes AVA ou ARR après avoir actionné la commande COM.

Un exemple : la session UG28.9, dont la synthèse mentionne 3 affichages, alors que 18 références ont effectivement été visualisées, dont 15 sous forme complète. A la suite d'une recherche sur le sujet *CD-ROM*, l'utilisateur a obtenu une liste de 14 références, dont il a voulu afficher

toutes les notices. Il a consulté la première sous forme abrégée, a actionné la commande COM, puis la commande AVA qui lui a permis de faire défiler les notices 2 à 14 directement sous forme complète : ces affichages n'ont pas été comptabilisés.

Il s'ensuit que certaines recherches, dont la synthèse mentionnait un affichage inférieur ou égal à deux, devraient peut-être figurer dans notre corpus, mais seul un examen détaillé des sessions permettrait de le déterminer. (Cette remarque permet peut-être d'expliquer pourquoi certains usagers "experts" en matière d'OPACs semblent se servir aussi rarement de la commande COM.)

3- Corpus et méthode d'étude

Avant de caractériser le corpus de sessions retenues, il n'est pas inutile de fixer la définition d'un certain nombre de termes et d'expressions que nous utiliserons couramment dans la suite de cette étude.

3-1 Quelques définitions

Nous appellerons "recherche" l'ensemble des opérations conduites par l'utilisateur autour d'un thème donné. Une session peut donc se composer d'une ou plusieurs recherches. La plupart en comportent une ou deux, voire trois; une seule session (UG07.18) enchaîne un dizaine de recherches indépendantes. La notion de "thème" ou "sujet" de recherche reste cependant vague et subjective, d'où une ambiguïté persistante dans certains cas. Pour fonder une typologie des recherches, nous avons retenu pour critère la plus ou moins grande correspondance sémantique entre les termes entrés par l'utilisateur et les formulations de sa question dans le pré- et le postquestionnaire. Une "étape de recherche" se compose de l'ensemble des démarches accomplies entre deux retours au sommaire.

Rappelons qu'il existe 6 "modes de recherche" possibles, que l'utilisateur sélectionne au niveau du sommaire:

- 1 - TIT - Recherche par titre
- 2 - AUT - Recherche par auteur
- 3 - A-T - Recherche auteur-titre
- 4 - SUJ - Recherche par sujet
- 5 - NUM - Recherche par cote, indice, éditeur, ISBN, ISSN...
- 6 - MOT - Recherche par mot(s) du titre, de l'auteur ou du sujet.

Nous ne retiendrons pas le septième choix du sommaire, LIM, comme mode de recherche, puisqu'il ne sert qu'à délimiter la partie du catalogue de la Médiathèque dans laquelle l'utilisateur souhaite conduire sa recherche selon un mode donné. Une étape de recherche débute toujours par le choix d'un mode de recherche et un même mode peut être sélectionné durant plusieurs étapes.

3-2 Typologie des recherches et des utilisateurs

Les 154 sessions du corpus se répartissent en cinq classes distinctes:

1) les sessions en une étape et un mode de recherche ; ce sont des recherches que nous pouvons qualifier d'"homogènes": 46 soit 30 %.

2) les sessions en plusieurs étapes, mais un seul mode de recherche : 39 soit 25 %.

les sessions en plusieurs étapes et plusieurs modes de recherche, parmi lesquelles on peut distinguer :

3) les sessions qui combinent deux modes de recherche : 31 soit 20 %.

4) les sessions qui combinent trois modes de recherche ou plus: 6 soit 4 %.

5) les sessions regroupant plusieurs recherches indépendantes : 26 soit 17 %.

6 sessions, soit 4 %, n'ont pas été classées en raison du caractère ambigu de leur formulation : il s'agit des sessions AC.17.14, ME09.13, ME18.11, UG10.10, UG11.8, UG16.7, pour lesquelles il s'avère difficile de déterminer si l'utilisateur a poursuivi le même thème de recherche ou s'il a combiné plusieurs thèmes.

Les sessions dans lesquelles l'usager n'a mis en oeuvre qu'un mode de recherche, à une ou plusieurs reprises, représentent donc 45 % du total.

Le mode dominant est incontestablement le mode sujet : 77 sessions font appel exclusivement à ce mode, tandis que 119 (77 %) contiennent au moins une recherche par sujet.

Le tableau suivant permet de situer le degré de familiarité avec l'OPAC des utilisateurs dont nous avons retenu la session. Ces données peuvent être confrontées à celles fournies pour l'ensemble de l'enquête dans le rapport d'étude qui lui a été consacré⁴⁰. (Nous avons regroupé certaines catégories).

	eff. corpus	eff. enquête
Usage < 1 fois par mois	: 78 50,5 %	305 51 %
Usage = 1-2 fois par mois	: 55 36 %	185 31 %
Usage > 1 fois par semaine	: 20 13 %	108 18 %

E. Kolmayer propose une caractérisation des usagers⁴¹ :

- l'"usager expert" se sert de l'OPAC de la médiathèque plus d'une fois par semaine.

- l'"usager novice" l'utilise moins d'une fois par mois.

La catégorie intermédiaire représentera pour nous celle des utilisateurs moyens.

40- Résultats de l'étude statistique. In [ECOLE 93], p.54.

41- Kolmayer, E., Analyse des échanges opérateur-système lors de l'interrogation de GEAC 9000. In [ECOLE 93], p. 76.

Notre corpus offre une proportion légèrement supérieure d'utilisateurs experts, mais stable de novices : l'ensemble de sessions retenu ne modifie pas de manière significative la répartition des utilisateurs et peut être considéré sur ce point comme représentatif de l'enquête.

3-3 Méthode d'étude

Nous nous proposons d'étudier dans un premier temps le passage de la question 24 du préquestionnaire à la première formulation entrée, c'est-à-dire la manière dont l'utilisateur formalise sa question au moment de commencer sa recherche et la transforme en requête lors du démarrage de la session proprement dite : nous comparerons pour cela PR24 avec le premier terme de recherche entré par l'utilisateur.

Nous nous intéresserons ensuite à l'évolution de cette formulation durant le déroulement même de la session, selon la typologie en 5 classes que nous avons dégagée. Nous étudierons en particulier les phénomènes d'interactivité qui s'instaurent - ou ne s'instaurent pas - avec le système et leur influence sur la formulation de la requête.

Dans son étude "Analyse des échanges opérateur-système lors de l'interrogation de GEAC 9000"⁴², réalisée sur un autre corpus de sessions enregistrées à la Médiathèque de La Villette, E. Kolmayer estime que l'usager utilise peu l'interactivité permise par le système : qu'il s'agisse des commandes de renvois, de la reprise de termes proposés par le système ou de l'instanciation d'une référence connue, c'est-à-dire la visualisation d'une notice complète et la reprise de ses termes d'indexation pour enrichir une requête. Elle note également, à juste titre, qu'une étude fine de cette interactivité nécessiterait de disposer d'un retour verbalisé de l'usager sur sa recherche : nous ne disposons pas des postquestionnaires verbaux enregistrés à l'issue de certaines sessions. Une étude systématique des échanges linguistiques entre l'usager et le système nous semble cependant suffisante pour confirmer ou infirmer certaines hypothèses émises en ce qui concerne l'interactivité.

Pour finir, nous étudierons le passage de la question PR24 et de la session à la question 11 du postquestionnaire. Cette confrontation ne sera cependant pas toujours possible, le nombre de sessions dotées d'une réponse à cette question P011 s'avérant particulièrement réduit : 39 sur 154, soit 25 %. Il semble que les utilisateurs, à l'issue de sessions particulièrement denses et après deux formulations libres du sujet, aient jugé superflu de le reformuler. Mais l'absence de réponse ou les remarques personnelles peuvent également constituer des retours intéressants sur la session.

42- Kolmayer, E., Id., p. 75-85.

II- Evolution de la formulation, du préquestionnaire à la première requête

3 usagers n'ont pas répondu à la question 24 du préquestionnaire, un quatrième s'est contenté d'une réponse vague : *ouvrages de cours*. Les sessions AC27.10, ME09.8, TTA04.4 et UG10.10 ne se prêtent donc à aucune comparaison.

Un autre ensemble de 6 sessions rend impossible toute confrontation entre la question PR24 et le premier terme de recherche entré par l'utilisateur, sinon pour constater qu'elles ne présentent aucun rapport.

Divers facteurs permettent d'expliquer ce phénomène :

- L'utilisateur n'a pas compris la question. Il a indiqué son objectif, et non son sujet de recherche : c'est le cas en UG10.9 (PR24 = *savoir travailler sur un ordinateur*).
- L'utilisateur a modifié son thème de recherche entre PR24 et le début de la session : ainsi dans ME12.3 (PR24 = *information sur l'organisation de chantier en génie civil / SUJ = aotocad*) ; ME27.7 (PR24 = *Oracle.../ SUJ = le langage Ada*) ;
- La session comprend plusieurs recherches successives ou intercalées et PR24 ne correspond pas nécessairement à la première : AC13.1, AC17.14 et UG28.9 en constituent des exemples.

Inversement, certaines réponses à la question PR24 regroupent plusieurs thèmes de recherche qui peuvent être

- soit indépendants : AC14.7 (PR24 = *information sur le métier de pâtissier de restau information sur des jeux en société*) ; TTA21.1 et TTA24.5 suivent ce modèle.
 - soit connexes mais distincts : du type AC27.5 (PR24 = *information sur le marché mondial du bois et des fleurs exotiques*) ; AC18.9, AC24.7, ME24.2, UG07.5, UG23.2, TTA09.3 et MHS20.1 entrent dans ce cas de figure.
- L'utilisateur sélectionne l'un de ces thèmes lors de sa première recherche.

1- Typologie des premières requêtes

Le tableau suivant offre une répartition par mode de recherche des 154 premières requêtes du corpus :

Mode de recherche	Nombre de sessions	Pourcentage
TIT	: 25	16,2
AUT	: 11	7,2
A-T	: 3	2
SUJ	: 104	67,5
NUM	: 1	0,6
MOT	: 10	6,5
Total	: 154	100,0

11 usagers experts de l'OPAC sur 20 (55 %) commencent de préférence leur recherche par une requête auteur ou titre, contre 36 % des utilisateurs moyens ; les novices commencent en majorité par une recherche sujet : 61 sur 78, soit 78 %. Le questionnaire d'enquête ne permet cependant pas de déterminer avec certitude si ce choix constitue une stratégie délibérée, ou s'il est lié au degré d'expertise du domaine de recherche.

Qu'appelle-t-on d'ailleurs "expert du domaine" ? Celui qui l'a étudié ou l'étudie, celui qui connaît des ouvrages dans ce domaine ? Il nous semble que ce degré ne pourrait être déterminé en toute certitude que lors d'un entretien avec l'utilisateur. Nous reviendrons ultérieurement sur ce point.

2- Comparaison PR24 / première requête par titre (25 cas)

Nous n'avons identifié qu'un seul cas de correspondance linguistique exacte entre la question PR24 et la première requête par titre : celui de la session ME20.3, *bases de données et systèmes relationnels*. Cette référence figurait d'ailleurs dans le catalogue de la Médiathèque, d'où son affichage immédiat par le système.

Dans 10 cas, le titre que l'utilisateur a proposé en première recherche se trouve inclus dans la formulation énoncée en PR24 :

- Le titre de livre ou de revue est cité en PR24, entre guillemets :
UG16.7 : PR24 = *L'ouvrage "A brief history of time" de Stephen Hawking en anglais.* (5 cas)
- Le sujet énoncé en PR24 inclut le titre qui sera proposé en première requête. 3 exemples : ME04.4 (PR24 = *logique et jeux mathématiques / TIT = jeu mathématique*), ME05.14, UG13.2.
- La première requête résulte d'une troncature du titre proposé en PR24, bien que la troncature soit peu utilisée dans l'ensemble des sessions que nous avons étudiées.
ME19.3 : PR24 = *avion de chasse / TIT = avion de.* (2 exemples)

13 premières requêtes ne présentent aucune correspondance linguistique, mais seulement un lien sémantique avec la question PR24, qui comporte un énoncé de sujet sans référence à un titre précis. L'objet de la recherche n'est alors plus nécessairement le titre lui-même.

TTA09.2 :
PR24 = *thérapie par le jeu*
TIT = *éducation à domicile de l'enfant IMC*
Sélection et affichage de la référence 2 *L'éducation à domicile de l'enfant infirme moteur cérébral.*

La dernière session, ME09.8, n'a pas reçu de réponse à la question PR24.

7 titres seulement ont immédiatement été reconnus par le système, 6 titres proches ont été sélectionnés dans la liste proposée par le système, tandis que 12 recherches ont échoué. Il est cependant permis de se demander si certains utilisateurs qui interrogent par titre connaissent effectivement le titre entré ou s'ils proposent en recherche par titre ce qui pourrait aussi constituer un sujet : en effet, 5 des utilisateurs ayant commencé leur session par une recherche par titre ont également affirmé dans le préquestionnaire ne pas connaître de livres ou de revues sur le même sujet ou le même auteur, donc être des novices dans le domaine de recherche.

3- Comparaison PR24 / première requête par auteur (11 cas)

Parmi ces 11 sessions, on relève 4 cas dans lesquels le nom de l'auteur figure aussi dans la question PR24 :

UG07.18 : PR24 = AUT = *Tolkien*

AC11.7 : PR24 = *science fiction (stephen king)*
AUT = *king stephen*

AC13.11 : PR24 = *livres de rené dumont*
AUT = *e dumon*

TTA18.3 : PR24 = *comptabilite analytique de C. RAULT*
AUT = *rault*

3 de ces recherches n'ont pas abouti (2 échecs au moins s'expliquent par une inadéquation entre la demande et le fonds de la Médiathèque de La Villette). L'auteur "René Dumont" a été sélectionné dans une liste d'auteurs lors d'une seconde tentative respectant la règle d'écriture proposée par le système, qui prévoit de commencer par le nom de l'auteur, et non par son prénom.

Dans les 7 autres cas, l'utilisateur a proposé à titre de première requête un nom d'auteur, jamais de prénom, en rapport direct avec le sujet proposé en PR24. 5 requêtes ont permis d'afficher une liste de noms dans laquelle l'usager a sélectionné l'auteur souhaité. Cette technique a échoué, du moins dans un premier temps, en ce qui concerne la requête AUT = *Jackobiak*, l'orthographe véritable du nom d'auteur étant "Jakobiak". Une requête, ME06.8, a débouché sur un affichage de deux noms d'auteurs, que l'utilisateur n'a pas retenus (à tort, semble-t-il, si l'on en croit la suite de la session).

Comme dans le cas des premières requêtes par titre, 3 utilisateurs ont estimé, dans le préquestionnaire d'enquête, qu'ils ne connaissaient aucun document sur leur sujet ou auteur, réponse qui peut paraître surprenante. Il semblerait donc que cette question ait été source d'ambiguïté pour certains utilisateurs.

4- Comparaison PR24 / première requête par auteur-titre (3 cas)

3 sessions affichent comme première requête une recherche par auteur-titre : AC23.4, UG10.9 et UG11.3. Nous avons déjà signalé que la première requête de UG10.9 ne présentait aucun rapport sémantique ou linguistique avec la question PR24. L'utilisateur, qui s'entraîne visiblement à utiliser l'OPAC, reprend d'ailleurs en titre une formulation proposée en exemple par le système.

Les 2 requêtes restantes reprennent soit une partie du titre, soit les éléments auteur-titre déjà mentionnés dans la question PR24 :

UG11.3 :

PR24 = *ouvrages sur la programmation en turbo pascal par ex.*

A-T = *f. madoule*
turbo pascal

AC23.4 :

PR24 = *conception et gestion des systèmes documentaire de van slype*

A-T = *van slype*

conception et gestion des syst

Ce faible échantillon ne permet pas une véritable étude de la recherche A-T, d'ailleurs rarement utilisée au cours des sessions enregistrées. Elle témoigne, davantage que la recherche par titre, d'une réelle connaissance d'un ouvrage concernant le domaine de la recherche.

Les premières requêtes par titre, auteur, auteur-titre pourraient servir de points de départ à la mise en oeuvre d'une stratégie d'instanciation, l'utilisateur reprenant dans la suite de sa recherche des termes d'indexation contenus dans les notices affichées en réponse. 40 sessions (26 % du corpus) seulement commencent par une recherche de ce type, c'est-à-dire par une entrée de données connues de l'utilisateur. Si une telle stratégie d'instanciation existe, elle ouvre rarement la recherche, alors que cela constituerait pourtant une de ses principales fonctions.

5- Comparaison PR24 / première requête par mot(s) du titre, de l'auteur ou du sujet (10 cas)

7 sessions commencent par une recherche sur mot(s) du titre, soit que le mode MOT ait été directement sélectionné, soit qu'il s'agisse d'une recherche en mode TIT transformée par la commande MOT.

La session ME07.4 porte sur 2 mots figurant dans la question PR24, "capteurs" et "pression" : par la syntaxe adoptée, *T:{capturs / pression}*, l'utilisateur choisit d'obtenir les références contenant l'un ou l'autre terme dans leur titre. On aurait pu s'attendre à ce qu'il croise les deux termes : il a cependant pu commettre une erreur de syntaxe (cette requête succède en effet à 3 essais refusés par le système). La recherche n'a pas donné les résultats escomptés par suite d'une faute de frappe dans le mot "capteurs", mais a permis d'afficher des références, notamment sous forme complète.

Les 6 autres recherches portent sur un terme simple, emprunté à PR24, avec 2 cas de passage du pluriel au singulier :

TTA19.5 : PR24 = *lasers* / MOT T = *laser*

ME28.6 : PR24 = *technologie des réseaux locaux* / MOT T = *réseau*

3 sessions présentent une première requête par mot du sujet, mot simple (*oracle*, *wordperfect*) ou nom-sigle (ODA). L'un d'entre eux, *oracle*, ne figurait pas dans la question PR24 de la session UG16.10 *informations sur sgbd et informatique en général* : l'utilisateur a spécifié sa question en passant du préquestionnaire à la recherche proprement dite.

Entre recherche par mot du titre ou du sujet, il s'avère parfois difficile de trancher. Il faut noter que la plupart des recherches par MOT présentent un grand nombre de refus du système de prendre en compte les termes entrés par les utilisateurs : ceux-ci ne respectent que très

rarement la syntaxe exigée pour ce mode de recherche (syntaxe assez complexe pour un utilisateur non habitué à pratiquer ce type de recherche, et présentée sur un écran particulièrement dense et peu convivial). Les cas de découragements sont fréquents, les utilisateurs revenant à une recherche par titre ou par sujet, ou essayant alternativement recherche par mot du titre et par mot du sujet : la session UG28.7 en constitue un bon exemple [2 tentatives S: ODA, puis 3 T: ODA, correctes mais interprétées comme "OD" par le système, avant de parvenir à T:(ODA) : on ne peut pas parler de véritable stratégie].

6- Comparaison PR24 / première requête par sujet (104 cas)

104 sessions commencent par une recherche sujet. Avant d'étudier les correspondances linguistiques et/ou sémantiques qui peuvent exister entre la question PR24 et ces premières requêtes, il n'est pas inutile de les caractériser sur le plan linguistique et sémantique.

6-1 Typologie syntaxique

Le modèle majoritairement repérable est celui du **syntagme nominal simple**, qui se présente sous la forme d'un uniterme, d'un sigle, ou d'un nom suivi d'un adjectif ou d'un nom propre. On en trouve 74 cas (71%), ainsi répartis :

- Unitermes : 45 (dont 4 noms propres et 2 mots composés)
- AC12.6 : *multimédia*
- Sigles : 3
- ME03.10 : *tgao*
- Nom + adjectif : 24
- ME03.11 : *système informatique*
- Nom + nom propre: 2
- ME20.2 : *méthode SD*

On peut identifier 20 occurrences (19 %) de **syntagmes nominaux complexes**, du type nom + syntagme prépositionnel (SP)

ou nom + SP + SP (2 cas)

ou nom + adjectif + SP (2 cas)

- AC13.7 : *traitement du signal*
- UG11.12 : *population en asie du sud-est*

3 occurrences contiennent une **coordination** de 2 termes :

- ME16.8 : *informatique et culture*

ou de 2 syntagmes :

- ME13.1 : *suites mathématiques et informatique*
- AC25.1 : *participation financière des salaris et motivation*

Ces 2 derniers cas présentent d'ailleurs une ambiguïté sémantique, due à la coordination par "et". Dans ME13.1, l'utilisateur cherche vraisemblablement à croiser les termes "suites mathématiques" et "informatique" pour en étudier le rapport. Mais il se crée une ambiguïté entre le nom et l'adjectif "informatique" : le singulier pourrait permettre ici de lever l'ambiguïté, bien que l'abondance des fautes de frappe et

d'orthographe relevées dans les questionnaires et les sessions rendent hasardeux le recours à ce critère. Il en est de même dans AC25.1, du fait du syntagme nominal anaphorique "motivation (des salariés)". Ces cas sont cependant rares par rapport à ceux que l'on peut relever dans la formulation en langage naturel ou quasi-naturel de PR24 ou P011.

On trouve quelques **juxtapositions** de termes :

- par ellision de préposition, ou de préposition suivie d'article :
 AC14.7 : *profession patissier* (ellision de "de")
 AC17.19 : *protection nature* (ellision de "de la")

- par simple apposition de noms ou de syntagmes nominaux, qui traduit une volonté de spécifier le premier terme ou syntagme, trop général, au moyen du second :
 TTA25.2 : *nutrition Afrique*
 ME19.6 : *informatique normalisation*
 UG05.8 : *soins infirmiers décret de compétence*

Coordination et juxtaposition constituent des tentatives de recherche booléenne de type ET dans un mode de recherche qui ne le permet pas. Les utilisateurs concernés sont des novices ou des usagers moyens de cet OPAC et des OPACs en général. Leurs formulations auraient pu être utilisées comme telles, en respectant la syntaxe proposée par le système, dans le cadre d'une recherche par mots du sujet. Dans le mode sujet, elles n'ont pas abouti, le système ne proposant qu'une liste de vedettes proches du premier terme entré, donc du premier nom ou syntagme. La tentative de spécification du sujet a échoué : l'utilisateur a soit sélectionné une vedette équivalente ou proche du premier terme ou syntagme, soit recommencé sa recherche en proposant une autre formulation :
 AC25.1 : sélection de la vedette "participation aux bénéfiques--salariés"
 ME19.6 : deuxième recherche par le sujet *informatique de gestion*, expression qui reprend la réponse à la question PR24.

2 premières requêtes ne peuvent être classées, un problème d'ordre technique ayant coupé l'enregistrement du début de la session ou perturbé la frappe du sujet (UG21.1, TTA24.4).

6-2 Typologie sémantique

La répartition des requêtes par sujet selon les grandes catégories classificatoires que nous avons retenues dans le premier chapitre permet de dégager les tendances suivantes :

Philosophie	: 1
Littérature	: 0
Arts	: 1
Hist. Géo.	: 5
Sc. sociales	: 27
Sciences	: 13
Techniques	: 16
Informatique	: 26
Télécommunications	: 5

10 sujets demeurent difficiles à classer : ainsi, par exemple, "bois" ou "maintenance". Seule la suite de la recherche conduit à les répartir en sciences (1) / techniques (2) / informatique (3) et sc. sociales (4).

On relève une prépondérance du domaine scientifique et informatique, qui va de pair avec la nature du fonds de la Médiathèque et avec la spécialité, le domaine d'étude indiqués par les utilisateurs : les informaticiens et techniciens sont nombreux. Mais un grand nombre de requêtes porte également sur des sujets de société ou des thèmes économiques, pour lesquels le fonds de la Médiathèque est moins riche, voire inadapté.

6-3 De PR24 à la première requête par sujet

Nous écarterons de cette étude 4 requêtes sans rapport ni linguistique, ni sémantique avec PR24, ainsi que 2 sessions qui ne disposent pas de réponse à cette question (AC13.1, UG28.9, ME27.7, ME12.3, AC27.10, TTA04.4).

Il était permis de penser que l'utilisateur, influencé par sa réponse à la question PR24, reprendrait tout ou partie de sa formulation dans sa première recherche par sujet : 15 premières requêtes conservent en effet l'intégralité de la réponse et la moitié des recherches (52 sur 104) en sont extraites.

En revanche, 17 recherches, sémantiquement liées à la question PR24, diffèrent radicalement sur le plan lexical et syntaxique et 14 requêtes ne reprennent qu'un terme ou une expression. La comparaison entre la question PR24 et la première requête par sujet rend alors possible une étude des modifications linguistiques apportées par l'utilisateur dans le but d'adapter son sujet de recherche à ce qu'il connaît ou croit avoir compris du système.

6-3-1 Identité (15)

On peut parler d'équivalence linguistique - syntaxique et sémantique - au sens strict dans 9 cas sur les 15 recensés.

AC13.7 :

PR24 = SUJ = *traitement du signal*

Il s'agit d'occurrences dans lesquelles la question PR24 comportait déjà une formulation concise du sujet de recherche, sous la forme d'un uniterme ou d'un syntagme nominal du type [N + A] ou [N + SP].

A ce chiffre s'ajoutent 3 occurrences dans lesquelles l'utilisateur a entré une formulation identique, mais en commettant une faute de frappe.

L'équivalence s'accompagne à 2 reprises d'un passage du pluriel au singulier. Il témoigne d'une certaine connaissance du système et des règles documentaires, qui exigent que les termes de recherche soient entrés au singulier.

AC25.5 :

PR24 = *tatouages* / SUJ = *tatouage*

La session TTA16.3 présente une modification de préposition ou plus exactement une perte de l'article qui figurait à l'état d'amalgame dans PR24 :

PR24 = *culture du champignon* / SUJ = *culture de champignon*

Cette conservation de la formulation peut s'expliquer par le fait que l'utilisateur n'a pas "joué le jeu", en renonçant à expliciter son sujet : les questions PR24 de ces sessions ne comportent pas de véritables descriptions de la recherche ultérieure.

6-3-2 Inclusion du sujet dans la réponse PR24 (52)

Il arrive que la question PR24 regroupe plusieurs thèmes de recherche et que la première requête ne soit que l'un d'entre eux.

AC27.5 :

PR24 = *informations sur le marché mondial du bois et des fleurs exotiques /*
SUJ = *bois*

Le passage à la première recherche s'opère alors par fractionnement du sujet global de recherche, simplification exigée par un système qui ne prend pas en compte les thèmes multiples. Il est cependant intéressant de constater qu'à 5 reprises les thèmes écartés de la première requête seront en définitive abandonnés par l'utilisateur : s'agit-il d'un oubli ou d'une lassitude due à une session trop longue à son gré ?

35 occurrences de PR24 comportent une indication de support, au sens large du terme, qui sera évidemment supprimée lors du passage à la première requête. Nous avons déjà mentionné qu'il pouvait s'agir d'une reprise des formulations citées en exemple.

Les termes introducteurs se répartissent comme suit :

information(s), infos	: 18
ouvrage(s), livre(s), documents, documentation	: 14
formulations, définitions	: 1
diapositives	: 1
texte de loi	: 1

Certains de ces termes ne sont pas vides sur le plan sémantique : "diapositives", "texte de loi". Leur élimination, lors du passage à la session, représente donc une perte d'information.

La formulation proposée en PR24 a parfois subi des transformations syntaxiques : il s'agit d'occurrences dans lesquelles des syntagmes nominaux complexes ont dû être adaptés au système et donc transformés en syntagmes nominaux simples. On relève 15 cas, et plusieurs types de transformations, souvent combinés :

- la perte d'un syntagme prépositionnel, qui spécifiait le sujet

ME14.3 :

PR24 = *la qualité dans la gestion de projet informatique /* SUJ = *qualité*

- inversement, un syntagme prépositionnel, de niveau 1 ou 2, devient centre de syntagme

UG19.4 :

PR24 = *la gestion de la culture /* SUJ = *culture*

- la perte d'une expansion, d'une apposition ou d'un syntagme coordonné

ME21.14 :

PR24 = *informations sur certains circuits logiques comme les bascules JK et D /* SUJ = *Circuits logiques*

- dans certains cas, la structure [N + A] est ramenée à N ou à A .

ME25.8 :

PR24 = *documentation sur les mémoires dynamiques et leur problème de rafraîchissement* / SUJ = *mémoire*

Dans chacun de ces cas, la simplification syntaxique induit un élargissement du sujet. Un tel élargissement est d'ailleurs rarement obtenu par troncature : la session TTA25.5 en présente bien un exemple, mais elle semble liée à la longueur du sujet plus qu'à la volonté de l'utilisateur.

On ne relève aucune tentative de construction syntaxique de vedette-sujet composée à partir des termes de la question PR24. Au contraire, la seule formulation proche d'une vedette composée, *base de données.méthodologie*, perd son extension. Nous avons mentionné en II-6-1 la prépondérance des syntagmes nominaux simples : l'utilisateur est suffisamment averti des contraintes du système pour simplifier son sujet de recherche, mais il ne connaît pas assez les règles documentaires ni les termes d'indexation retenus par le système pour "construire" véritablement une vedette-sujet. Ce constat est valable quels que soient la formation de l'utilisateur et son degré de familiarité avec l'OPAC.

Les seuls indices d'une connaissance du système restent :

- le passage du pluriel au singulier, observé dans 9 cas sur 10.
 - l'abandon des prédéterminants, à 2 exceptions près : TTA10.4 et TTA20.6.
 - l'abandon des prépositions à l'intérieur d'un syntagme : TTA25.2.
- Cet inventaire a de quoi décevoir.

6-3-3 Abandon de la formulation initiale (17)

Dans plus de la moitié des cas, l'utilisateur a visiblement jugé son sujet trop spécifique pour être proposé comme tel : il en a abandonné la formulation initiale pour le remplacer par un terme générique ou une expression moins technique.

ME07.6 :

PR24 = *modulation de fréquence* / SUJ = *télécommunication*

UG26.1 :

PR24 = *info sur tests fungus sur alliages aluminium + info sur coeff friction* / SUJ = *moisissure*

Ce passage du spécifique au générique ne se justifie pas toujours, car il oblige ensuite à restreindre la recherche ; il constitue néanmoins une stratégie de départ chez bon nombre d'utilisateurs. Une étude des réponses au préquestionnaire d'enquête permet cependant de constater que cette stratégie n'est pas le fait d'experts de l'OPAC, puisque 8 des utilisateurs peuvent être considérés comme des novices, le dernier se présentant comme un utilisateur moyen.

Plus rarement, on observe au contraire une spécification d'un sujet trop général : 3 cas.

TTA26.3 :

PR24 = *documentation sur les logiciels* / SUJ = *reflex*

Ces requêtes portent sur des sujets techniques ou informatiques.

A 4 reprises, la reformulation du sujet s'accompagne d'un déplacement, que les données en notre possession ne permettent pas toujours d'expliquer. Ainsi, la session UG05.8 :

PR24 = *information sur les mst chez les adolescents / SUJ = soins infirmiers decret de compéten/*

Les 2 formulations du sujet sont vraisemblablement liées dans l'esprit de l'utilisateur, comme le montre la réponse apportée par l'un d'eux à la question 11 du postquestionnaire : *quelle est le role de l'infirmiere dans la prévention des mst.*

L'utilisateur, peu ou pas habitué à cet OPAC, ne parvient pas à "fixer" la formulation de son sujet.

6-3-4 Un terme ou un syntagme commun (14)

On relève dans ce groupe quelques tendances déjà évoquées précédemment :

- le passage d'un syntagme nominal complexe à un syntagme simple, du type N ou [N + A]. L'utilisateur condense sa formulation, mais cherche à la rendre aussi équivalente que possible, sur le plan sémantique, à PR24, d'où la reprise de certains termes.

AC24.7 :

PR24 = *techniques électromagnétiques en géophysique ou recherche pétrolière / SUJ = électromagnétisme géophysique*

- élargissement de 2 sujets, jugés trop spécifiques

UG11.12 :

PR24 = *croissance démographique en asie du sud est /SUJ = population en asie du sud est*

- 2 cas, inverses, de spécification du sujet

TTA16.5 :

PR24 = *livres sur les dessins d'enfants / dessin d'enfants handicapés*

3 sessions font état cependant d'un effort de l'utilisateur pour combiner 2 termes de recherche :

- soit en les juxtaposant

AC14.7 :

PR24 = *information sur le métier de patissier de restau / SUJ = profession patissier*

- soit en les coordonnant par "et"

ME16.8 :

PR24 = *conséquences du traitement automatique du langage naturel sur la langue et la culture / SUJ = informatique et culture.*

Synthèse

La majorité des premières requêtes s'effectue donc en mode sujet. Ces requêtes peuvent être analysées sous plusieurs angles :

- **l'angle syntaxique** : on note une prépondérance des syntagmes nominaux simples (74 sur 104) et complexes (20). Les quelques tentatives de coordination et de juxtaposition de syntagmes constituent des essais de recherche booléennes, alors que le mode SUJ ne s'y prête pas.

- **l'angle sémantique** : il révèle une prépondérance - attendue - du domaine scientifique et technique, mais également la part importante des sujets relatifs aux sciences sociales, pour lesquels la recherche s'avèrera nécessairement moins riche.

Les **premières requêtes par sujet** présentent une forte ressemblance linguistique avec la question 24 du préquestionnaire :

- 15 sont totalement identiques et témoignent du fait que l'utilisateur ne veut pas ou ne peut pas reformuler son sujet. - 52 découlent de PR24 par suppression de termes le plus souvent vides et empruntés aux exemples du préquestionnaire, ou par transformations syntaxiques du type SN complexe -- > SN simple.

- 14 présentent un lien syntaxique faible, par un terme : l'utilisateur a cherché à résumer sa réponse à PR24, en recombinaut certains mots.

- 17 premières requêtes ne conservent qu'un lien sémantique avec PR24. L'utilisateur éprouve des difficultés à déterminer à quel niveau sémantique il doit lancer sa recherche : elle est fonction de son domaine de recherche et de la représentation de ce domaine dans le langage de l'OPAC, représentation dont les utilisateurs occasionnels ignorent tout.

Les **premières requêtes par titre et par auteur** présentent soit un lien sémantique avec PR24 (13 et 4), l'utilisateur ayant formulé un sujet dans le préquestionnaire, soit une reprise le plus souvent partielle de la réponse (11 et 7), montrant qu'il centre sa recherche sur un titre ou un auteur précis.

Les **premières requêtes par mot** reprennent pour la plupart (9 sur 10) un élément de la réponse à PR24, élément recherché dans le titre (7) ou le sujet (2).

Les **premières requêtes par auteur-titre ou par cote** sont trop peu nombreuses pour permettre des conclusions.

On observe en définitive que l'utilisateur auquel on a au préalable demandé de formuler sa question en langue naturelle ne se détache plus ou à grand peine de cette formulation : 93 sessions présentent un lien syntaxique (reprise ou inclusion) et sémantique avec PR24, 38 un lien sémantique, qui n'est jamais un lien de synonymie ; 14 recomposent certains termes particulièrement significatifs de PR24.

Nous pouvons nous demander si l'utilisateur fait ensuite évoluer la formulation de son sujet au cours de la session, ou s'il se laisse guider par le système, ce qui tendrait à prouver qu'il n'est pas capable d'effectuer lui-même cette opération. Une étude des choix linguistiques de l'utilisateur en cours de session devrait apporter un début de réponse à cette question.

III- Etude de la formulation durant le déroulement de la session

1- Problèmes d'utilisation des commandes

L'examen des sessions permet tout d'abord de mettre en évidence un certain nombre de problèmes liés à l'utilisation des commandes de l'OPAC. Ces difficultés sont le fait de certaines commandes, souvent les mêmes, dont les utilisateurs ne parviennent pas à distinguer les fonctions. Les sessions que nous avons étudiées révèlent notamment une confusion quasi-générale entre les instructions LIS, CHO, et AVA/ARR, perceptible dans les recherches par sujet.

Prenons l'exemple d'une recherche relativement "simple", telle que ME03.4. L'utilisateur a conduit une recherche par sujet, en 2 étapes : après une tentative infructueuse au moyen de l'expression *communication parallèle*, il a élargi sa formulation à *télécommunication* (écran n°6). Il a choisi une des 10 vedettes sujets proposées par le système à l'écran n°7 : "microordinateur--télécommunication", pour laquelle le système a proposé une liste de références auteur/titre comprenant 3 réponses (écran n°8). Les 8 écrans qui suivent montrent que l'objectif de l'utilisateur fut alors de visualiser les 3 notices correspondantes. Pour y parvenir, il a mis en place une stratégie :

Ecran 8 : sélection de la référence n°1

E.9 : notice 1 sous forme abrégée - sélection de la commande LIS (= retour à la liste)

E.10 : cette commande renvoie à la liste des vedettes sujets, donc à l'écran n°7. L'utilisateur sélectionne de nouveau la vedette sujet "microordinateur--télécommunication".

E.11 : il est ramené à l'écran des 3 références correspondantes, écran n°8. Il sélectionne la seconde référence.

E.12 : notice 2 abrégée - sélection de la commande LIS.

E.13 = E.10 = E.7 : liste des vedettes sujets. L'utilisateur sélectionne pour la troisième fois la vedette "microordinateur--télécommunication".

E.14 = E.11 = E.8 : il sélectionne la référence n°3.

E.15 : notice 3 abrégée - sélection de la commande SOM (= retour au sommaire).

E.16 : sommaire (et fin de la session).

Cette stratégie révèle 2 types de lacunes :

- tout d'abord une méconnaissance de la commande CHO, très rarement utilisée, qui permet de rappeler la liste des références liée à une vedette sujet, et non pas la liste - antérieure - des vedettes sujets. L'utilisateur, qui n'a pas toujours eu le temps ou le réflexe de mémoriser les titres intéressants pour sa recherche, devrait se reporter à ces références pour effectuer une nouvelle sélection de notice. Dans les faits, il n'en est rien. L'utilisateur actionne spontanément la commande LIS.

Il nous semble que cette attitude a une origine linguistique. Pour l'utilisateur, "LIS = retour à la liste" doit permettre de revenir à la liste qui l'intéresse en priorité, celle des références abrégées auteur/titre. D'autre part, la commande CHO, par sa dénomination même, est ambiguë : "choix de références" donne probablement à l'utilisateur l'impression qu'il s'agit d'un nouveau choix à effectuer alors qu'il n'a

pas choisi ses références : il a choisi un sujet, pour lequel le système lui a proposé des références. Ce type de confusion entre les commandes CHO et LIS apparaît dans un minimum de 45 sessions sur les 154 étudiées, soit 30 % du corpus !

Un commentaire d'utilisateur, relevé dans le postquestionnaire correspondant à la session ME20.8, en donne confirmation. Voici sa réponse à la question n°7 : *Expliquez pourquoi vous avez trouvé cette recherche difficile ou impossible :*

Parfois les ordres ne sont pas assez clairs. On se retrouve alors sur des menus non demandés. Ex. quand on consulte une fiche dans une liste, peut-on revenir sur la liste des fiches sans passer par la liste des sujets ?

Pour cet utilisateur, et pour beaucoup d'autres, il existe deux 'listes', celle des sujets et celle des fiches, autrement dit des références. Au cours de sa recherche, LIS l'a renvoyé à la première, alors qu'il souhaitait revenir à la seconde.

Un apprentissage des fonctions distinctes de ces commandes, l'adjonction d'une explication à l'écran ou l'adoption d'une nouvelle dénomination pour la commande CHO permettraient d'alléger considérablement certaines recherches plus complexes que ME03.4. Dans la session que nous venons d'analyser, l'utilisation de CHO aurait permis d'éliminer 2 écrans.

- Un grand nombre d'utilisateurs ne maîtrisent pas non plus la technique du feuilletage et ses commandes : AVA/ARR. Ainsi, dans la session précédente, il suffisait à l'utilisateur d'actionner la commande AVA à partir de la première notice abrégée (écran n°9) pour passer à la seconde notice, puis à la troisième : soit une recherche totalisant 12 écrans distincts, au lieu de 16. Cette recherche était simple et le nombre de notices à visualiser suffisamment réduit pour autoriser la stratégie mise en place par l'utilisateur. Mais elle s'avère très lourde dans certains cas.

Les commandes AVA/ARR sont proposées à l'utilisateur avec des définitions en apparence explicites : notice suivante/notice précédente. Mais existe-t-il, pour des utilisateurs qui se repèrent mal dans le système, une notice suivante ou précédente par rapport à celle qu'ils sont en train de visualiser ? Elles existent bien dans la liste de références, puisqu'elles leur sont proposées sur un même écran. Mais on peut penser que certains utilisateurs perdent tous leurs repères lors du passage à la consultation des notices : "notice suivante" et "notice précédente" désignent alors pour eux "suite" ou "début" de notice.

Sur le plan linguistique, ces deux types de problèmes perturbent la formulation de la requête : d'une part, ils allongent inutilement la recherche, ce qui pousse souvent l'utilisateur à interrompre la session sans avoir exploité toutes les possibilités offertes par le système. D'autre part, ils obligent l'utilisateur à multiplier les interactions avec l'OPAC - à sélectionner plusieurs fois la même vedette ou à entrer les mêmes termes - ce qui constitue une source d'erreurs importante.

2- Recherches en une étape et un mode de recherche

46 sessions se composent d'une unique recherche, en une étape et selon un mode : l'utilisateur a entré au départ un terme ou une expression et s'en est tenu à cette formulation. C'est le type même de recherche que l'on pourrait qualifier de "pauvre", puisqu'elle ne comporte aucune reformulation du sujet, ni extension, ni restriction sémantique. 21 personnes se sont pourtant déclarées satisfaites de leur recherche, 7 la considérant comme un échec total (ils ont répondu "rien du tout" ou "peu de chose" à la question 1 du postquestionnaire). Ce modèle de recherche n'est pourtant pas le fait d'utilisateurs inexpérimentés, puisque 5 experts et 18 utilisateurs moyens l'ont adopté.

2-1 Caractérisation des recherches

La plupart d'entre elles s'effectuent selon le mode sujet:

- sujet : 36
- titre : 5
- auteur : 2
- mot : 3

Nous avons comptabilisé et étudié dans ce groupe 5 sessions qui comportent en fait plusieurs étapes, mais des étapes identiques : l'utilisateur a recommencé sa recherche de base à plusieurs reprises, par suite d'un problème technique ou d'une erreur d'utilisation des commandes. Sa démarche reste cependant celle d'une recherche en une étape.

Il s'agit de sessions relativement courtes, comprenant de 10 à 43 écrans, en moyenne 19. Après l'entrée par l'utilisateur de son terme de recherche, elles évoluent soit par sélections multiples dans une liste de titres, de noms d'auteur ou de sujets proposée par le système, soit par sélection unique et affichage de références. On constate que, dans 16 sessions, l'utilisateur n'a pas fait usage des commandes de feuilletage (AVA, ARR) ou les a utilisées sans succès, par suite de problèmes techniques ; 11 sessions comportent exclusivement des feuilletages de listes, 4 exclusivement des feuilletages de notices, et 12 combinent les deux. Le feuilletage des termes proposés par le système n'intervient donc que dans la moitié des cas. L'utilisateur se contente, dans l'autre moitié, des propositions de départ du système.

Les stratégies de recherche nous semblent en définitive pouvoir être classées en 6 modèles de base, dont 3 principaux.

2-2 Modèles de recherche

Modèles 1 : l'utilisateur sélectionne une vedette de type uniterme ou vedette composée, qui inclut son terme de requête ou lui correspond strictement ; sur le plan syntaxique, le terme entré peut figurer en tête de vedette ou en subdivision. Il peut être traduit dans la syntaxe du système.

On observe alors 3 cas de figure :

1-A - il retourne à la liste de vedettes en cours de session pour sélectionner d'autres propositions du système (de 1 à 8) : 17 cas. Il s'agit cependant toujours de vedettes incluant son terme de recherche.

1-B - il retourne à la liste mais sélectionne exclusivement la même vedette : 9 cas.

1-C - il affiche des références sans revenir à la liste : 4.

Exemple-type d'un modèle 1-A : ME28.5

PR24 : *information sur les techniques de maintenance des ordinateurs à distance*

Terme entré : *maintenance*

Réponse : au moins 10 sujets

Choix E.4 : "maintenance assistée par ordinateur"

Choix E.6 : "maintenance assistée par ordinateur--étude de marché"

Choix E.8 : "maintenance gestion voir gestion de la maintenance"

Choix E.11 : "maintenance gestion voir gestion de la maintenance"

Choix E.22 : "microordinateur IBM--maintenance"

Choix E.26 : "logiciel--maintenance"

Modèles 2 : le système ne reconnaît pas le terme entré ; l'utilisateur sélectionne alors une vedette proche. Il ne peut s'agir que d'une vedette dont la tête reprend le premier terme entré ou une partie de ce terme. On retrouve les 3 cas précédents :

2-A - retour à la liste et choix d'autres vedettes : 11 cas.

2-B - retour à la liste et sélection de la même vedette : 1.

2-C - affichage de références : 1 cas.

Exemple-type d'un modèle 2-A : TTA16.6

PR24 = *ENFANCE MALTRAITEE*

Terme entré : *enfance maltraitée*

Réponse : 0

Choix écran 6 : "enfant"

Choix écran 13 : "hospitalisation--enfant"

Les modèles de base 1-A et 2-A dominent : l'utilisateur exploite alors, plus ou moins, les propositions du système. On ne compte cependant que 13 cas (30 %) de modèles 2, c'est-à-dire de reformulations (au sens de traduction dans le langage de l'OPAC) et/ou déplacements du sujet par le système. Les sessions en une étape et un mode de recherche sont donc essentiellement des sessions dont le sujet est nettement défini et conforme au langage de l'OPAC.

2-3 Stratégies de recherche

A 14 reprises, l'utilisateur n'a sélectionné qu'une vedette titre, auteur ou sujet : toute sa recherche consiste à exploiter cette unique sélection d'un terme proposé par le système.

Soit la session AC27.9 (18 écrans) :

PR24 : *ouvrages récents sur les politiques énergétiques de pays du tiers monde*

Terme de recherche entré par l'utilisateur (E.4) : *pays en voie de développement energ*

Réponse du système (E.5) : 0

Feuilletages de la liste de sujets proches : 5 ARR

Choix d'un sujet proposé par le système (E.10) : "énergie--économie--pays en voie de développement"

Liste de 6 références (E.11)

Affichage des 6 références (écrans 12 à 18)

Le nombre moyen de vedettes sélectionnées sur proposition du système est de 2,9 par session, chiffre plutôt faible, et de 2,3 si on ne relève que les vedettes différentes : une fois sur deux, l'utilisateur sélectionne à plusieurs reprises la même vedette. Cela s'explique soit par les problèmes de maîtrise des commandes que nous avons déjà évoqués, soit par une absence de méthode dans la sélection des vedettes ; l'utilisateur peut avoir oublié qu'il a déjà choisi un sujet.

Les 3 recherches par mot (ME07.4, UG13.6 et TTA18.8) sont à distinguer, car elles donnent toutes lieu à de nombreux affichages de notices à partir d'un unique sujet de recherche (13, 15 et 30) : l'utilisateur examine la totalité des références proposées par le système.

Dans 63 % des cas, l'utilisateur a proposé un titre, un nom d'auteur ou un sujet figurant dans l'index de l'OPAC ;

dans 7 % des cas, le système a fourni un titre, un nom ou un sujet proche sur le plan syntaxique (par exemple TTA16.3 - terme entré : *culture de champignon* / terme proposé par le système : "culture des champignons voir champignon : culture");

30 % des termes entrés ne figuraient pas dans l'index, ce qui a conduit l'utilisateur à choisir un terme proche (par exemple TTA20.6 - terme entré *le syndicalisme en europe* / termes sélectionnés sur proposition du système: "syndicalisme--politique", "syndicat", "syndicat--europe").

3- Recherches en plusieurs étapes et un mode de recherche

Ce modèle de recherche concerne 39 sessions du corpus. Il semble à priori plus riche que le modèle précédent sur le plan du dialogue usager/OPAC, puisque l'utilisateur, tout en revenant au même mode de recherche (par sujet, titre, auteur ou mot), a entré de nouveaux termes, qui peuvent éventuellement refléter la mise en place de stratégies linguistiques et induire des déplacements sémantiques. Ce sont les deux aspects auxquels nous nous intéresserons, après avoir caractérisé les recherches.

Les auteurs de ces recherches présentent le degré d'expertise suivant par rapport au système :

- novices : 24
- moyens : 11
- experts : 4

La proportion d'utilisateurs novices est donc plus forte dans ce type de recherche que dans le type précédent (61,5 contre 50 %).

3-1 Caractérisation des recherches

La répartition par mode s'effectue comme suit :

- Sujet : 30
- Titre : 5
- Auteur : 1
- Mot : 3

L'une des sessions (ME07.6) aurait pu être classée avec les recherches en plusieurs étapes et 2 modes, mais l'utilisateur n'est pas parvenu à formuler sa recherche par mot, et s'en est finalement tenu à une recherche par sujet.

Il s'agit de recherches plus longues que les précédentes : entre 14 et 84 écrans, 29 en moyenne. Les 3 recherches par mot sont les plus longues: 52, 83 et 84 écrans, l'utilisateur multipliant les entrées de termes. Le nombre moyen d'étapes, et donc de termes de recherche entrés, est de 3,7 ; les recherches par mot atteignent cependant une moyenne de 10 étapes.

Chacune des étapes réunit un nombre moyen de 8 écrans, chiffre nettement inférieur au nombre moyen d'écrans des recherches en une étape (19). Plus les étapes se multiplient, plus elles sont courtes, ce qui peut, dans certains cas, avoir une influence sur leur performance : l'utilisateur revient au sommaire et modifie sa recherche sans avoir épuisé les possibilités offertes par le système. Le chiffre suivant en apporte un début de preuve : l'utilisateur choisit en moyenne 4 propositions du système, mais seulement 3,5 vedettes différentes. Il revient donc en cours de recherche à des propositions trop vite abandonnées, sans qu'il ne s'agisse, le plus souvent, d'une stratégie de sa part.

3-2 Modèles de recherche

Définir des modèles communs à des recherches complexes s'avère difficile, voire impossible : chaque session présente une configuration qui lui est propre, comme l'indique le tableau de synthèse suivant. Il est cependant possible de montrer comment s'articulent les modèles que nous avons définis précédemment (cf. paragraphe III-2-3). Chaque session enchaîne en effet plusieurs modèles de base. Prenons l'exemple de la session ME03.11 :

PR24 = *information sur les système informatiques*

Etape 1 : SUJ

Terme entré : *système informatique*

Réponse : 9 sujets

Choix E.4 : "système informatique"

E.6 : affichage d'une notice

Etape 2 : SUJ

Terme entré : *unix*

Réponse : 3 sujets

Choix E.10 : "unix voir système d'exploitation unix"

E.12 : affichage d'une notice

Choix E.13 : "unix voir système d'exploitation unix"

E.15 : affichage d'une notice

La première étape de cette session correspond au modèle 1-C (terme de recherche inclus dans la vedette choisie / affichage de références sans retour à la liste), la seconde au modèle 1-B (terme de recherche inclus dans la vedette choisie / retour à la liste et sélection de la même vedette). On peut donc modéliser l'interaction linguistique de l'utilisateur avec le système sous la forme : 1-C/1-B.

Il nous faut cependant ajouter 2 modèles de base, que la nature même du corpus (minimum de 3 notices affichées) rendait inexistantes dans le cas de recherches en une étape. Il s'agit des étapes dans lesquelles l'utilisateur entre un terme de recherche, et retourne au sommaire sans avoir sélectionné de vedette ou après avoir sélectionné une vedette mais aucune notice. Nous les appellerons 1-D ou 2-D, selon que le système reconnaît ou non le terme entré par l'utilisateur. Ces étapes sont "vides" dans la majorité des cas : l'utilisateur n'est pas satisfait des vedettes proposées par le système et il abandonne le terme entré. (Ce n'est cependant pas toujours vrai : certains utilisateurs peuvent très bien se contenter d'afficher des références très abrégées, sous forme auteur/titre.)

26 sessions contiennent un ou plusieurs modèles 1-D ou 2-D, dont on dénombre au total 59 occurrences (mais les 3 recherches par mot en totalisent 23 à elles seules). Si on excepte ces recherches par mot, 29 sur 36 n'ont donné lieu à aucune sélection de vedette. Connaître la place de ces modèles dans la session peut alors s'avérer important dans la mesure où ces étapes ont probablement obligé l'utilisateur à reformuler son sujet ou à le déplacer sur le plan sémantique.

A titre d'exemple, le tableau suivant propose une synthèse, session par session, des modèles repérés :

AC13.7 : 1-C / 1-C / 2-C
 AC13.11 : 2-D / 2-A
 AC16.6 : 2-D / 2-D / 1-B / 2-D
 AC16.9 : 2-D / 1-A
 AC18.9 : 1-A / 2-D / 1-C
 AC24.7 : 2-D / 1-A / 1-D / 1-A
 AC25.5 : 1-C / 2-C / 2-D / 1-C
 ME03.4 : 2-D / 1-C
 ME03.11 : 1-C / 1-B
 ME05.14 : 1-A / 1-C
 ME07.6 : 1-A / 1-C / 1-A / 1-C / 1-C / 2-C / 2-D
 ME09.8 : 1-C / 1-C
 ME12.1 : 1-A / 1-B
 ME12.3 : 2-D / 1-A / 1-A
 ME12.11 : 1-C / 2-C / 1-C / 1-C
 ME16.5 : 1-A / 1-C / 2-C
 ME16.8 : 2-D / 2-D / 1-A / 2-D
 ME17.10 : 2-D / 1-A
 ME19.4 : 2-D / 2-D / 2-D / 1-A
 ME20.2 : 2-B / 2-B / 2-D
 ME21.14 : 2-D / 1-A
 UG03.7 : 1-C / 1-D / 2-A / 1-A
 UG10.4 : 1-A / 2-D / 1-C
 UG11.12 : 2-C / 1-C / 1-C
 UG12.8 : 1-C / 1-C / 1-D / 2-D
 UG14.3 : 1-C / 2-C / 1-C
 UG16.10 : 1-D / 1-D / 1-D / 1-D / 2-D / 2-D / 1-C / 1-D / 1-C [MOT]
 UG16.13 : 2-D / 1-A / 1-C
 UG18.3 : 2-D / 1-B
 UG19.2 : 1-C / 1-C / 1-B
 UG19.4 : 1-D / 1-C / 1-A / 2-D / 2-D
 UG20.3 : 1-D / 2-D / 2-D / 2-D / 2-D / 2-D / 2-D / 2-D / 1-C
 1-C / 1-C / 2-D / 2-D / 2-D / 2-D / 1-D / 1-D [MOT]
 TTA04.5 : 1-A / 2-C / 2-A
 TTA10.4 : 2-D / 1-D / 1-A
 TTA11.5 : 1-A / 2-C / 2-D
 TTA13.2 : 2-C / 1-B
 TTA16.5 : 2-C / 1-A
 TTA19.5 : 1-C / 1-D / 1-C / 1-C / 1-C / 1-D / 1-C [MOT]
 TTA24.4 : 1-A / 2-D / 1-C / 2-D / 1-B / 2-D

Le modèle de base 1-D apparaît peu en dehors des recherches par mot, mais sa présence est révélatrice : il témoigne du fait que le terme entré par l'utilisateur n'est pas adapté à la recherche en cours, car trop générique ou ambigu sur le plan sémantique : ainsi le terme "culture" entré dans le cadre d'une recherche sur la gestion de la culture (session UG19.4) est pris dans son sens agricole par le système. En AC24.7, UG03.7, UG19.4 et TTA10.4, l'usager choisit ensuite d'aborder le sujet par un autre de ses aspects. En UG12.8, il précise son sujet, sans succès.

Une comparaison entre les modèles de type 1 et les modèles de type 2 permet dans une certaine mesure d'évaluer qui, de l'utilisateur ou du

système, dirige la formulation du sujet. E. Kolmayer distingue 2 cas de figure :

L'utilisateur a une démarche dirigée par les concepts lorsqu'il propose des termes d'interrogation à l'OPAC ; sa démarche est dirigée par les données lorsque l'OPAC lui propose des listes de termes parmi lesquels il choisit⁴³.

Dans les sessions dominées par des modèles de type 1, les concepts entrés par l'utilisateur sont repris par le système : c'est le cas par exemple de la session UG19.2, où les termes *ariane*, *satellite* et *navette spatiale* figurent dans l'index sujet.

Au contraire, dans les sessions dominées par les modèles de type 2, on observe un écart entre le langage utilisé par l'utilisateur et celui du système : 8 sessions sont partiellement ou totalement composées de modèles de ce type. L'OPAC dirige alors en grande partie la formulation et l'évolution de la question. Exemple : la session ME20.2, dans laquelle les termes *méthode SD*, *interface utilisateur* et *sd* ne sont pas reconnus par le système, qui propose "méthode SADT" et "interface--programmation". Cet écart linguistique n'est observé ici que dans des sessions d'utilisateurs moyens (2) ou novices (6).

3-3 Evolution sémantique

On peut distinguer 4 types fondamentaux d'évolution sémantique des formulations :

- la restriction du domaine sémantique du sujet, autrement dit le passage d'un niveau générique à un niveau spécifique d'interrogation.
- l'extension du domaine sémantique, ou passage d'un niveau spécifique à un niveau générique.
- le déplacement sémantique, c'est-à-dire le changement de domaine d'interrogation ou l'association de plusieurs domaines ; mesurer ce déplacement s'avère souvent délicat, voire subjectif, et le recours à l'utilisateur serait souvent nécessaire pour expliciter son cheminement.
- la stabilité sémantique, rare mais attestée dans certaines sessions.

Certaines sessions combinent 2 types d'évolution, rarement davantage.

Seules les recherches par mode sujet sont concernées, ce qui conduit à éliminer les sessions suivantes :

- AC13.11 : recherche qui porte sur un auteur
- TTA19.5 : recherche par mot du titre, uniquement sur le mot *laser/lasers*.
- les recherches par titre ME05.14, ME09.8, ME12.11, ME16.5 et ME17.10.

On observe 10 cas d'extension du domaine d'interrogation, contre 4 restrictions. Il semble que l'utilisateur ait davantage tendance à interroger d'emblée à un niveau spécifique, puis, n'obtenant aucune réponse satisfaisante, à élargir sa formulation.

43- Kolmayer, E., Analyse des échanges opérateur-système lors de l'interrogation de GEAC 9000. In [ECOLE 93], p. 76, note 1.

Extension : ME03.4
 PR24 : *protocole centronics (liaison parallele)*
 1er terme : *communication parallèle*
 Réponse : 0
 2ème terme : *télécommunication*

Restriction : ME03.11
 PR24 : *information sur les systeme informatiques*
 1er terme : *système informatique*
 Réponse : 9 sujets
 2ème terme : *unix*

L'utilisateur peu familiarisé avec les OPACs peut aussi alterner extensions et réductions du domaine d'interrogation ; c'est le cas dans 5 sessions d'utilisateurs novices AC16.6, ME20.2, UG20.3, UG16.10 et surtout ME07.6 (le niveau d'interrogation, G pour générique et S pour spécifique, est indiqué à la suite) :

PR24 : *modulation de fréquence*
 1er terme : *telecommunication (G)*
 Réponse : au moins 10 sujets
 2ème terme : *modulation (G --> S)*
 3ème terme : *telecommunication (S --> G)*
 4ème terme : *téléphone (G --> S)*
 5ème terme : *communication (S --> G)*
 6ème terme : *telecomunication (S --> G)*
 7ème terme : *électronique telecommunication (stable)*

14 sessions ont engendré des déplacements de sujet : il est parfois nécessaire de recourir à la question 24 du préquestionnaire et à la question 11 du postquestionnaire pour les justifier.

La session AC16.9 présente un déplacement à priori inexplicable :

1er terme : *logitiel*
 2ème terme : *commodore*

La question PR24 montre cependant que les deux sujets sont liés dans l'esprit de l'utilisateur : *logitiel pour commodore amiga*.

Le déplacement correspond en fait à un éclatement du sujet de recherche initial en 2 thèmes ou points d'accès possibles. On observe la même démarche dans les sessions AC18.9 et TTA10.4, AC24.7, UG03.7, UG11.12, UG12.8, UG19.2, UG19.4 et TTA24.4.

Une variante de cette démarche est un déplacement progressif, comme le montre la session AC25.5 :

PR24 : *tatouages*
 1er terme : *tatouage*
 2ème terme : *tatouages polynésiens*
 3ème terme : *polynésie française*
 4ème terme : *tahiti*

Le second terme de recherche introduit une délimitation géographique, qui devient ensuite thème principal de recherche, l'utilisateur tentant d'approcher le sujet sous ce nouvel angle. Le cas est unique dans ce groupe de sessions.

Troisième cas de figure : un déplacement que ni la formulation initiale du sujet ni les réponses apportées par le système ne permettent

d'expliquer et dont on mesure mal comment il pourrait permettre à l'utilisateur de mieux cerner son sujet.

Ainsi la session ME16.8 :

PR24 : *consequences du traitement automatique du langage naturel sur la langue et la culture*

1er terme : *informatique et culture*

2ème terme : *intelligence artificielle*

3ème terme : *informatique* (passage à un terme générique)

Les sessions UG10.4 et TTA04.5 suivent ce modèle. Les 3 recherches ont été menées par des utilisateurs novices et non spécialistes du domaine d'interrogation.

3-4 Instanciation

Aucune des recherches par auteur ou par titre de ce groupe de sessions n'a donné lieu à un affichage de notice complète susceptible de fournir de nouveaux termes de recherche : le fait que la recherche se soit poursuivie sur le même mode en constitue d'ailleurs une preuve.

Une des 3 recherches par mot contient 7 notices complètes, mais il s'agit d'une session dans laquelle l'utilisateur n'a entré que 2 termes de recherche différents, à 7 reprises : *laser* et *lasers*. Il n'a donc pas utilisé l'indexation des notices complètes.

La session ME05.14 présente peut-être un cas de reprise de données fournies par le système, mais il ne s'agit que d'une hypothèse. D'après la question PR24 et sa première requête, l'utilisateur recherche des documents sur excel, pas sur windows. Il utilise le mode titre. Après avoir affiché la notice "Excel 3 pour windows" (écran 8), il recommence sa recherche avec le titre *windows*. Il est vraisemblable qu'il a été influencé par la recherche précédente et les titres fournis par le système.

Il semble en définitive que les recherches en plusieurs étapes et un mode ne se prêtent guère aux techniques élaborées de reformulation. On ne peut cependant nier l'existence d'une interactivité avec le système, mais une interactivité "par défaut" : devant les refus du système de reconnaître ses termes de recherche, l'utilisateur se doit de réagir et de proposer non pas des reformulations linguistiques, mais des déplacements sémantiques. Cette interactivité débouche cependant sur des tâtonnements, rarement sur une recherche cohérente.

4- Recherches en plusieurs étapes et plusieurs modes de recherche

On distinguera 2 types de sessions, selon le nombre de modes utilisés:

- les sessions comprenant une recherche en plusieurs étapes et 2 modes de recherche : 31 cas
- les sessions comprenant une recherche en plusieurs étapes et un nombre de modes supérieur ou égal à 3 : 6 cas, dans lesquels l'utilisateur a utilisé 3 ou 4 modes, jamais davantage.

La session ME25.6 présente 3 modes de recherche (TIT, SUJ, MOT), mais elle a été classée dans les sessions en 2 modes, l'utilisateur n'étant pas parvenu à faire fonctionner la recherche par mot.

Après avoir caractérisé ces occurrences, nous tenterons de décrire l'influence que peut avoir le changement de mode sur la formulation linguistique de la requête : pourquoi l'utilisateur abandonne-t-il un mode pour un autre ? Comment choisit-il ce nouveau mode de recherche ? Reformule-t-il alors sa question ou conserve-t-il sa formulation initiale ? On peut difficilement tenter d'étudier l'évolution sémantique de la formulation dans le cas des recherches en modes multiples, étant donné que le mode sujet n'est pas nécessairement choisi ou conservé. Mais, comme précédemment, nous étudierons les rares cas d'instanciation qui pourraient être dégagés dans ces sessions.

4-1 Caractérisation des recherches

Les sessions en plusieurs modes sont de loin les plus longues : les sessions en 2 modes affichent une moyenne de 41 écrans et de 4,5 étapes, les sessions en 3 ou 4 modes une moyenne de 64 écrans et de 10,5 étapes. Le nombre moyen d'écrans par étape est donc respectivement de 9 et de 6.

Le niveau d'expertise par rapport au système varie cependant : les experts et utilisateurs moyens dominent dans les sessions en 2 modes (7 experts, 10 utilisateurs moyens, soit 55 %), alors que les sessions en 3 modes sont menées par des utilisateurs moyens (5) et par un novice. La multiplication des modes de recherche ne serait donc pas le signe d'une plus grande maîtrise du système, mais au contraire d'une certaine dispersion dans la stratégie de recherche, alors que la combinaison judicieuse de 2 modes peut apparaître comme une stratégie.

Les modes de recherche associés dans les sessions en mode double se répartissent comme suit :

TIT / SUJ : 11
 AUT / SUJ : 4
 MOT / SUJ : 8
 MOT / TIT : 3
 AUT / TIT : 2
 A-T / SUJ : 1
 NUM / TIT : 1
 NUM / SUJ : 1

Le mode sujet peut être associé à tous les autres, comme le confirme la répartition des sessions en 3 ou 4 modes :

AUT / TIT / SUJ : 2
 TIT / MOT / SUJ : 2
 AUT / TIT / MOT du sujet : 1
 AUT / TIT / A-T / SUJ : 1

Les experts conduisent en priorité une recherche selon les modes AUT ou tit et SUJ (4), mais aussi MOT / SUJ, MOT / TIT et NUM / TIT. Ils utilisent donc davantage les connaissances qu'ils peuvent avoir sur le domaine pour enrichir leur recherche et ne s'appuient pas uniquement sur les propositions du système. Nous verrons cependant qu'ils ne vont pas jusqu'à mettre en place de véritables stratégies d'instanciation, du moins dans notre sélection. Une fois encore, le nombre d'utilisateurs experts est trop limité pour en tirer des conclusions : une telle recherche devrait être conduite sur un nombre plus large d'utilisateurs.

On distingue 2 types de sessions en plusieurs modes :

- les sessions dans lesquelles les modes sont mis en oeuvre successivement : 24 cas.

ME28.2

1ère étape : SUJ = *suites de fibonacci*
 2ème étape : SUJ = *combinatoire*
 3ème étape : SUJ = *fibonacci*
 4ème étape : TIT = *fibonacci*

- les sessions dans lesquelles les modes sont alternés : 13 cas, dont 4 dans des sessions à 3 ou 4 modes.

UG07.5

1ère étape : AUT = *cuvier*
 2ème étape : SUJ = *géologie physique des antilles*
 3ème étape : AUT = *laserre guy*
 4ème étape : SUJ = *stratigraphie de la guadeloupe*

L'utilisateur qui mène une session en mode double sélectionne en moyenne 5,25 propositions du système, soit un peu plus d'une par étape, chiffre comparable à celui des recherches en plusieurs étapes et un mode de recherche, mais nettement inférieur à celui des recherches en une étape et un mode (2,9). Il en choisit 9,5 dans une session à 3 ou 4 modes. Il entre donc davantage de termes, mais exploite moins les propositions du système. Il serait difficile d'expliquer l'origine de ce phénomène sans interroger chaque utilisateur.

Nous pouvons cependant signaler qu'à 3 exceptions près, la réponse de l'utilisateur à la question 24 du préquestionnaire contenait des syntagmes nominaux complexes ou des thèmes connexes coordonnés ou juxtaposés, c'est-à-dire une description du sujet qui se prêtait peu à l'élaboration d'une requête concise : d'où la difficulté de devoir passer du thème à la requête puis reformuler le sujet en cas d'insatisfaction. Le recours à un mode différent pourrait-il être fonctionner comme un palliatif ?

4-2 Mode de recherche et évolution de la formulation

Il nous semble inutile de reproduire ici les différents modèles de base que l'on peut identifier dans ces sessions : comme dans le cas des recherches en plusieurs étapes et un mode de recherche, chaque session possède sa propre configuration. Certaines recherches en mode double présentent cependant un trait commun : dans 18 cas sur 31, un changement de mode intervient après un modèle 2-D, autrement dit après l'entrée d'un terme non reconnu par l'OPAC et des propositions du système refusées par l'utilisateur. On retrouve ce cas de figure dans 3 sessions en 3 ou 4 modes.

AC11.7

1ère étape : AUT = *king stephen*

Réponse : 0 (pas de sélection de vedette proche)

2ème étape : SUJ = *science fiction*

L'échec d'un mode de recherche justifie bien, dans ces conditions, l'essai d'un autre mode. On peut alors se demander ce qu'il advient de la formulation du sujet : l'utilisateur la conserve-t-il d'un mode à l'autre ? Dans le cas des sessions en mode double, 10 utilisateurs ont effectivement repris tout ou partie d'une formulation antérieure, 6 ont tenté une nouvelle formulation, 2 ont exploité un autre terme de la question 24 du préquestionnaire.

Le phénomène est moins apparent dans les recherches multiples, qui procèdent davantage par tâtonnements et reprise d'un terme dans le même mode, parfois avec le mode LIM. La session UG21.10 offre un parfait exemple de conservation de la formulation d'un mode à l'autre, dans le cadre d'une recherche en 2 modes :

PR24 = *les puzzles*

1ère étape : SUJ = *puzzle*

Réponse : 0 (pas de sélection de sujet proche)

2ème étape : TIT = *puzzle*

Plus généralement, on dénombre 7 cas de conservation de la formulation d'un mode à un autre sur les 18 sessions de type AUT/SUJ, TIT/SUJ, A-T/SUJ. Le phénomène est plus systématique dans le cas des sessions faisant intervenir le mode MOT : il y a reprise de termes dans 10 cas sur 11, soit un total de 17 cas.

ME17.15

PR24 = *information sur les bases de données objet ainsi que sur les méthodes de développement associées*

1ère étape : TIT = *hood*

Réponse : 0

2ème étape : MOT - T : *hood*

Réponse : 2

3ème étape : MOT - T : *versan*

Réponse : 0

4ème étape : MOT - T : *parallèle*

Réponse : 5

5 sessions en 3 ou 4 modes contiennent aussi des reprises, mais jamais sur 3 modes : d'AUT à AUT/TIT, de SUJ à TIT, de SUJ ou TIT à MOT.

La principale utilisation de la recherche par mot consiste donc à reprendre ou à tester des termes qui ont déjà été proposés avec peu ou pas de succès dans une étape précédente ou le seront dans la suivante. Cette stratégie semble éviter une reformulation à l'utilisateur suffisamment expérimenté pour la mettre en oeuvre. (Il en utilise rarement les possibilités, en particulier la combinaison de termes). Au lieu de reformuler, d'élargir ou de préciser le sujet, il reprend certains termes, parfois non significatifs : ainsi, dans la session AC13.4, l'utilisateur conduit une recherche par mot sur les termes "avantages" et "comparaison", et interroge sur ce dernier terme dans le mode sujet, sans succès !

Or, la recherche par mot a pour but d'élargir la recherche, mais ne dispense pas d'une reformulation du sujet. Mise en oeuvre sur des termes qui ne sont pas adaptés au langage du système, elle amène davantage de bruit que de références pertinentes : d'après les modèles de base que nous avons proposés (en particulier un repérage des modèles de type 1), seules 5 recherches sur 11 semblent avoir fourni des résultats, satisfaisants ou non pour l'utilisateur.

4-3 Instanciation

Nous avons prêté une attention particulière aux sessions dans lesquelles l'utilisateur a visualisé une ou plusieurs notices complètes, donc avec indexation, dans un mode AUT, TIT ou A-T. Seules 2 sessions pouvaient avoir donné lieu à une reformulation au moyen de termes fournis par le système : UG07.5 et UG16.9. Dans le premier cas, il n'existe aucun rapport entre la notice et le terme de recherche suivant, qui figurait déjà dans la question PR24. Dans le second, la notice complète est une référence liée, qui ne présente pas d'indexation.

En revanche, les 2 sessions suivantes témoignent peut-être d'un réinvestissement d'informations fournies par le système :

- ME19.6 : après avoir affiché, dans une recherche sujet, 2 notices abrégées possédant une cote du type I 6..., l'utilisateur a conduit une recherche en mode NUM, par la cote *i* 6. Il a sélectionné par ce biais une nouvelle référence dans son domaine de recherche, l'informatique de gestion.

- ME10.3 : à la suite d'une recherche sujet sur le terme *refrigeration*, l'utilisateur n'a obtenu qu'une seule référence, celle d'un ouvrage de Michel Guyot. La dernière étape de la session tente une recherche par auteur sur le nom *guyot*, qui ramène d'ailleurs à la même notice.

Ces 2 occurrences sont les seuls cas d'instanciation que nous avons identifiés pour les sessions utilisant 2 modes de recherche. Aucun n'a pu être repéré dans les sessions en 3 ou 4 modes.

5- Sessions composées de recherches indépendantes

26 utilisateurs ont conduit, au cours d'une même session, plusieurs recherches indépendantes, le plus souvent successivement (20 cas), parfois aussi en les intercalant (6 cas). Ces dernières sessions se caractérisent souvent par leur manque de méthode, l'utilisateur menant à plusieurs reprises la même recherche, proposant les mêmes termes et sélectionnant les mêmes notices.

ME11.5 :

PR24 = *information sur haut-parleur et micro*

1ère étape (rech.1) : SUJ = *haut-parleur*

2ème étape (rech.1) : SUJ = *microphone*

3ème étape (rech.1) : SUJ = *acoustique*

4ème étape (rech.1) : SUJ = *haut-parleur*

5ème étape (rech.2) : SUJ = *engrais*

6ème étape (rech.1) : SUJ = *haut-haut-parleur* (faute de frappe)

7ème étape (rech.1) : SUJ = *haut-parleur* (faute de frappe)

8ème étape (rech.1) : SUJ = *haut-parleur*

9ème étape (rech.1) : SUJ = *oreille* (faute de frappe)

10ème étape (rech.1) : SUJ = *oreille*

11ème étape (rech.2) : SUJ = *engrais*

12ème étape (rech.1) : SUJ = *haut-parleur*

Ces enregistrements comprennent en moyenne 35 écrans, ce qui est peu en regard du nombre moyen de recherches conduites : 3 (11 écrans par recherche). Il s'agit de recherches rapides, le plus souvent sans reformulation ou déplacement du sujet ; les possibilités du système sont peu utilisées. On dénombre 42 % de modèles de base 1-D ou 2-D, c'est-à-dire de refus par l'utilisateur des propositions du système et d'abandon de l'étape en cours (58 sur 139 modèles de base).

La répartition des modes par session s'effectue de la manière suivante :

- 19 sessions intègrent au moins une recherche en mode sujet,
- 12 une recherche en mode titre,
- 9 une recherche en mode auteur,
- 5 une recherche en mode mot,
- 2 une recherche en mode auteur/titre.

Les recherches se rattachent à l'un des types précédemment étudiés, mais leur intégration dans des sessions plus larges modifient leur évolution linguistique. Nous avons repéré au total 75 recherches distinctes:

- 46 recherches en une étape et un mode de recherche
- 18 recherches en plusieurs étapes et un mode de recherche
- 11 recherches en plusieurs étapes et plusieurs modes (2 ou 3).

5-1 Recherches en une étape et un mode de recherche (46 cas)

Le mode sujet n'est plus aussi dominant que dans le cas des sessions exclusivement composées d'une recherche de ce type (cf. III-2-1) : 48 contre 78 %. Les recherches en mode TIT sont également nombreuses :

- SUJ : 22
- TIT : 15
- AUT : 5
- A-T : 3
- MOT : 1

Ce type de recherche a une autre fonction que lorsqu'il est utilisé seul, celle de contrôler qu'un ouvrage existe bien dans le fonds de la Médiathèque. La preuve en est le pourcentage de ces recherches qui n'avaient pas été annoncées dans la question 24 du préquestionnaire : 29 sur 46, soit 63 %. Elles ne donnent pas lieu à un réinvestissement des données au cours d'une recherche ultérieure, puisqu'elles sont totalement indépendantes.

Sur le plan linguistique, les recherches qui se trouvaient mentionnées dans PR24 en reprennent systématiquement tout ou partie, à une exception près (AC27.8). L'utilisateur exploite ensuite les propositions du système : on repère 9 modèles de type -A ou -B ; mais plus encore, il se limite à un affichage de références : 24 occurrences des modèles 1-C ou 2-C sur 46 modèles. 28 % des recherches n'aboutissent pas. Dans 56 % des cas, l'utilisateur n'a sélectionné qu'une proposition du système :

- pas de sélection : 11
- 1 vedette sélectionnée : 26
- 2 vedettes distinctes : 5
- 3 vedettes distinctes : 2

Ces chiffres confirment le caractère rapide des recherches et leur pauvreté linguistique.

5-2 Recherches en plusieurs étapes et un mode de recherche (18)

On retrouve une forte proportion de recherches par sujet, comme dans le cas des sessions uniquement composées de ce type de recherche :

- SUJ : 12
- TIT : 2
- AUT : 2
- MOT : 2

8 de ces recherches n'ont pas été annoncées dans le préquestionnaire, les autres reprennent un thème de PR24, à l'exception de TTA04.4 qui ne possède pas de réponse à cette question.

Sur le plan syntaxique, la plupart des recherches présentent une conservation de termes d'une étape à une autre (12 cas). Ces recherches sont donc redondantes, sans que l'on puisse toujours déterminer la raison qui pousse l'utilisateur à proposer la même requête. Il arrive que l'utilisateur revienne à une recherche antérieure après en avoir commencé une autre. Sur les 4 recherches dissociées, 3 sont en effet des reprises.

AC13.1

PR24 : *logiciels micro-informatique*

1ère recherche :

SUJ = *système nerveux*

Réponse : au moins 9 sujets

Choix : 9 - "système nerveux central"

Réponse : 5 références

Affichage de référence : 1 - Braillon, Georges

2ème recherche :

TIT = *wordperfect*

TIT = *Word5*

TIT = *word 5*

1ère recherche :

SUJ = *système nerveux*

Réponse : au moins 9 sujets

Choix : 9 - "système nerveux central"

Réponse : 5 références

Affichage de référence : 5 - Morin, Georges

3ème recherche :

SUJ = *musique*

Une étude sémantique montre que les recherches que nous avons qualifiées de "stables" sont plus nombreuses : 6 sur 14 recherches par sujet ou par mot.

Cette répartition montre que l'utilisateur n'a pas pour priorité de faire évoluer son sujet, mais de conduire une recherche rapide, par essais et erreurs. La présence de nombreux modèles de type -D (24 sur 55 modèles recensés) le confirme.

5-3 Recherches en plusieurs étapes et plusieurs modes de recherche (11)

Il s'agit de recherches en 2 modes (8 cas) ou en 3 modes (3 cas) :

- AUT / SUJ : 2
- AUT / TIT : 3
- MOT / SUJ : 2
- AUT / A-T : 1
- AUT / TIT / SUJ : 1
- AUT / TIT / A-T : 1
- MOT / TIT / SUJ : 1

7 de ces recherches présentent une conservation de termes d'un mode à l'autre, totale (UG07.18) ou partielle (AC23.4 -2 cas-, AC27.8, ME05.11, TTA24.5, MHS20.1). L'objectif est, comme précédemment, d'étendre la recherche sans reformuler le sujet. Cette stratégie comporte parfois, là encore, un risque d'erreurs : ainsi, on note 2 conservations de termes du mode AUT au mode SUJ, l'utilisateur confondant peut-être une recherche sur des documents écrits par un auteur avec une recherche sur des ouvrages dont il est le sujet.

AC27.8

PR24 : *livres de gestion de papin*

1ère recherche :

SUJ = *association*

2ème recherche :

SUJ = *malaparte*

Réponse : 0

Pas de sélection de vedette proche

AUT = *malaparte*

Réponse : 0

Pas de sélection de vedette proche

Nous n'avons repéré aucune instanciation de données fournies par le système.

IV- Synthèse

1- Tableau récapitulatif

Le tableau suivant fournit une répartition des sessions par type de requête et par classe.

Rappelons que nous avons défini 5 classes :

- 1) les sessions en une étape et un mode de recherche
- 2) les sessions en plusieurs étapes mais un seul mode de recherche
- les sessions en plusieurs étapes et plusieurs modes de recherche, parmi lesquelles on distingue :
 - 3) les sessions qui combinent 2 modes de recherche
 - 4) les sessions qui combinent 3 modes de recherche ou plus
- 5) les sessions regroupant plusieurs recherches indépendantes.

lère requête	G.1	G.2	G.3	G.4	G.5	ambig.	total
TIT	5	5	7	1	3	4	25
AUT	2	1	3	1	4	0	11
A-T	0	0	1	0	2	0	3
MOT	3	3	3	0	1	0	10
NUM	0	0	1	0	0	0	1
SUJ	36	30	16	4	16	2	104
Total	46	39	31	6	26	6	154

2- Quelques stratégies de recherche

Sur 46 sessions en une étape et un mode de recherche, 14 ont donné lieu à la sélection d'une seule vedette auteur, titre ou sujet. Ne maîtrisant pas suffisamment les commandes de l'OPAC, l'utilisateur est fréquemment amené à sélectionner plusieurs fois la même vedette. Dans l'ensemble cependant, ces sessions sont relativement courtes et pauvres sur le plan de la reformulation parce que l'utilisateur propose un sujet concis dont le système accepte les termes dans 63 % des cas. Ces sessions se répartissent en 6 modèles élémentaires (1-A, 1-B, 1-C, 2-A, 2-B, 2-C).

Les 39 sessions en plusieurs étapes et un mode de recherche combinent 8 modèles de base : nous avons dû ajouter les modèles 1-D et 2-D, destinés aux étapes sans affichage de vedette ou de référence. Nous avons montré que, dans 22 cas, l'utilisateur dirigeait la formulation, mais avec des tâtonnements ; les sessions restantes révèlent un écart linguistique trop important entre le langage de l'utilisateur et celui du système, qui conduit à recommencer la recherche. On observe alors diverses évolutions sémantiques, notamment l'extension et le déplacement du sujet. L'utilisateur reformule rarement sa question de départ et il utilise

essentiellement les propositions du système, mais en lecture seulement, sans jamais les intégrer à sa requête.

Les 37 recherches en plusieurs étapes et plusieurs modes sont essentiellement des associations de 2 modes (31 sur 37). Nous avons remarqué que le changement de mode intervenait souvent après l'échec d'une requête, et qu'il pouvait alors consister en une entrée de la même requête dans un autre mode que celui dans lequel elle venait d'échouer. Le couplage de la recherche par mot(s) du titre ou du sujet à un autre mode de recherche et cette conservation des termes sont 2 stratégies majeures pour éviter de reformuler seul son sujet. Elles se révèlent cependant insuffisantes.

26 sessions composées de recherches indépendantes confirment un certain manque de méthode et de stratégies : la technique dominante est ici celle des essais et erreurs.

L'étude des 154 sessions de notre corpus confirme en définitive certaines hypothèses relatives aux difficultés linguistiques des utilisateurs d'OPACs et à la pauvreté de la reformulation. Une confrontation entre les réponses aux questions PR24 et P011 et les sessions de recherche s'impose cependant pour valider ces conclusions : nous l'avons menée sur les 39 sessions disposant d'une réponse à la question P011.

3- Question P011 et reformulation

Une session nous paraît illustrer ce que l'on pourrait appeler un cas de reformulation de la question : ME16.8.

PR24 = *consequences du traitement automatique du langage naturel sur la langue et la culture*

Recherche en plusieurs étapes et un mode :

SUJ = II

SUJ = *intelligence artificielle*

Sélection de vedettes proposées par le système :

"intelligence artificielle--philosophie"

"sciences cognitives--intelligence artificielle"

SUJ = *informatique*

P011 = *les enjeux humains et culturels du traitement automatique du langage naturel*

Indépendamment de la recherche qu'il vient de mener, l'utilisateur propose ici un nouvel énoncé, complet, de son sujet. Ce n'est pas le cas dans la plupart des autres sessions possédant une réponse à la question P011. Elles restent très proches de PR24 et de la session, montrant que l'utilisateur parvient difficilement à se détacher de la formulation première de sa question.

3-1 Identité ou inclusion

3 réponses à la question P011 sont strictement équivalentes sur le plan linguistique à la question PR24 : ME03.4 ; ME07.6 ; TTA09.6.

Exemple : ME03.4

PR24 = *protocole centronics (liaison parallele)*

P011 = *liaison parallele protocole centronics*

3 réponses sont incluses dans la réponse PR24 : AC12.6 ; AC17.19 ; ME28.5.

Exemple : ME28.5

PR24 = *information sur les techniques de maintenance des ordinateurs à distance*

P011 = *maintenance des ordinateurs à distance*

Inversement, 5 réponses PR24 ont été reprises et intégrées à la réponse P011 : AC11.7 ; AC13.1 ; UG13.3 ; UG14.3 ; TTA18.3. Les ajouts sont en majorité des termes désignant la nature du support souhaité : *roman, livres, ouvrages...*

La session UG14.3 associe au sujet de départ une restriction intéressante :

PR24 = *guide géographique*

P011 = *recherche d'un guide géographique (region paris.)*

1 réponse reprend une expression centrale de PR24 : ME21.14.

PR24 = *information sur certains circuits logiques comme les bascules JK et D*

P011 = *fonctionnement et table logique des bascules JK et D*

Ces 12 occurrences ne présentent pas de reformulation du sujet de départ. Un autre groupe de sessions comporte un essai de reformulation qui prend en compte le déroulement de la session précédente, qu'il s'agisse de termes entrés ou de réponses du système.

3-2 Reprise de termes

Ce groupe se compose de 22 sessions. 7 empruntent des termes à la session qui précède. Cette dernière peut donc avoir induit un déplacement du sujet, une spécification, une démarche, que l'utilisateur retient comme formulation "idéale" : AC27.9 ; ME11.7 ; ME13.1 ; ME24.2 ; UG16.10 ; TTA13.2 ; TTA20.10.

TTA20.10 :

PR24 = *information sur l'unification européenne*

Recherche en plusieurs étapes et plusieurs modes :

TIT = *tout maastricht*

SUJ = *europe*

SUJ = *europееurope*

SUJ = *europe*

SUJ = *concurrence*

P011 = *information sur la concurrence industrielle et sur la réglementation de la concurrence en europe*

L'influence d'une session décevante peut se manifester par un passage du spécifique au générique entre PR24 et P011 : on ne peut cependant pas parler d'une véritable reformulation du sujet. Ainsi, dans la session UG12.8, on passe des *musees d'etat de la ville de HUE au VIETNAM* aux *musees*

étrange, l'utilisateur ayant compris lors de sa recherche que les propositions doivent être plus vastes (c'est ce qu'il déclare dans la réponse 8 du postquestionnaire).

14 réponses font des emprunts à la réponse PR24 et à la session : AC16.6; AC16.12 ; ME04.4 ; UG04.12 ; UG05.8 ; UG09.2 ; UG16.8 ; UG21.8; UG26.5 ; UG28.8 ; TTA04.5 ; TTA11.5 ; TTA24.4 ; TTA26.3.

Ces reprises peuvent être interprétées de diverses manières :

- Certaines ne présentent pas de reformulation du sujet: il y a stabilité sémantique entre PR24, les termes entrés dans la session et P011. C'est le cas, par exemple, dans la session UG16.8 et 4 autres occurrences :

PR24 = *diapositives couleurs pour illustrer un article sur l'observatoire des télécommunication dans la ville*

Recherche en une étape et un mode :

SUJ = *tlcomunication dans la ville*

P011 = *je cherche des diapositives illustrant le sujet TELECOMUNICATION DANS LA VILLE*

- Les reprises de propositions du système, adjonctions de domaine ou d'aspect, développements du sujet en fonction de la session pourraient témoigner, quant à eux, d'une amorce de reformulation "assistée" par le système.

Ainsi TTA26.3 :

PR24 = *documentation sur les logiciels*

Recherche en une étape et un mode

SUJ = *reflex*

Le système propose "reflex voir logiciel reflex"

P011 = *logiciel reflex*

ou AC16.6 :

PR24 = *AUTOMATES PROGRAMMABLE*

Recherche en plusieurs étapes et un mode :

SUJ = *automates programmables*

SUJ = *telemecanique*

SUJ = *automate*

SUJ = *automates telemecanique*

P011 = *AUTOMATE PROGRAMMABLE TELEMECANIQUE*

3-3 Absence de lien

2 réponses à la question P011 doivent être éliminées, car elles ne présentent aucun rapport avec PR24 : il s'agit des sessions AC23.4 et UG21.1. La session TTA04.4 ne dispose pas de réponse à cette question.

Une seule session présente une réponse P011 sans lien syntaxique avec PR24, mais portant sur le même domaine : UG20.10.

PR24 = *information sur la bacteriologie medicale*

P011 = *différents ouvrages sur les pathologies*

Il s'agit d'ailleurs d'une session qui comprend plusieurs recherches indépendantes, dont la première seulement porte sur le sujet énoncé dans PR24.

On ne peut parler de reformulation, ni de passage du spécifique au générique, étant donnée la fragilité du lien sémantique qui unit les deux réponses.

Conclusion

Au terme de cette comparaison, nous observons que les quelques évolutions de formulation (17) sont essentiellement dues à la session précédente, alors que l'objectif de la question P011 était d'obtenir une formulation en langue naturelle, dégagée des contraintes du système : cette question P011, ainsi placée et formulée, n'atteint pas son objectif de départ. Mais elle permet de mettre en évidence que l'utilisateur reste, dans la plupart des cas, fortement dépendant du système pour la formulation et l'évolution linguistique de sa requête. Cette évolution ne peut être que freinée par la méconnaissance de l'OPAC et de son langage, méconnaissance qui est pourtant le lot de l'utilisateur moyen.

Cette étude des sessions d'utilisateurs d'OPAC nous a permis de définir des stratégies d'interrogation (recherches en une/plusieurs étapes, en un/plusieurs modes...) et des comportements linguistiques, mais il semble difficile de les interpréter sans faire appel à l'utilisateur. Seul l'auteur d'une recherche pourrait expliquer l'origine de tel ou tel choix linguistique, qui peut nous sembler inattendu. Ces données n'ont malheureusement pas été récoltées durant l'enquête de Juin 92.

Sans même disposer de ces informations, nous pouvons malgré tout nous interroger sur la pauvreté des stratégies linguistiques mises à jour.

Nous avons vu que l'on pouvait rarement parler de "reformulation", l'utilisateur se trouvant le plus souvent devant l'alternative suivante : le système a accepté sa requête, l'a éventuellement "traduite" dans son langage (sous la forme d'une vedette composée par exemple) ou bien il l'a refusée. Dans ce cas, l'usager recourt soit au feuilletage de listes de vedettes, soit au changement de mode de recherche (notamment sujet/titre, sujet/mot ou titre/mot), soit à un déplacement sémantique du sujet qui le laissera à terme insatisfait. L'utilisateur se trouve seul face à un thème qu'il connaît souvent mal, puisqu'il recherche des informations à son sujet.

Il serait cependant intéressant de savoir si l'utilisateur qui consulte un catalogue papier met en oeuvre des stratégies linguistiques plus élaborées, telles que l'utilisation systématique des synonymes ou des renvois : à notre connaissance, aucune enquête n'a encore été conduite pour comparer les stratégies mises en oeuvre, du moins en France.

Nous n'avons relevé que 3 occurrences potentielles d'une stratégie d'instanciation : l'utilisateur associe rarement ses connaissances et les notices complètes pour enrichir sa formulation de départ. Seules quelques réponses P011 témoignent d'une reprise de termes fournis par le système.

Cette absence n'est pas tout à fait surprenante : la plupart de nos études de sessions mettent en évidence un rapport d'immédiateté qui s'instaure entre l'utilisateur et la machine. L'utilisateur "zappe" d'une proposition du système à l'autre sans prendre toujours la peine d'exploiter une réponse, sans mémoriser le résultat de sa recherche pour le réinvestir, à quelques rares exceptions. Son type de recherche préféré est la recherche en une étape et un mode. Or, toute stratégie d'instanciation se déroule en deux temps : une recherche d'informations, puis la recherche proprement dite. Il est probable qu'elle exige trop d'anticipation pour des utilisateurs occasionnels et peu formés aux techniques documentaires. Le fait qu'elle n'apparaisse pas davantage chez les utilisateurs experts du

corpus peut surprendre davantage : ce résultat demanderait une confirmation sur un échantillon plus large d'utilisateurs formés aux OPACs.

Ces conclusions, à priori décevantes, confirment donc l'inadéquation qui existe entre le langage de l'utilisateur et celui du système. Mais elles fournissent aussi quelques pistes pour une poursuite - indispensable, voire urgente - des recherches dans ce domaine.

CHAPITRE 3

Linguistique et protocoles expérimentaux

L'enquête réalisée à la Médiathèque de La Villette a permis de collecter, pour la première fois en France, un corpus de données analysables sur le plan linguistique. Elle constitue en cela une étape nécessaire dans la recherche sur les OPACs. Il s'agit cependant, de l'avis même de ses concepteurs, d'une recherche *exploratoire* : la mise en évidence de ses résultats, mais aussi de ses faiblesses ou incomplétudes, devrait donc permettre de définir des orientations utiles pour une expérimentation future.

Des études similaires ou proches par leurs objectifs sont depuis longtemps conduites aux Etats-Unis et au Canada, notamment à la Northwestern University Library, l'Université de Wisconsin-Madison, l'Université de Californie Los Angeles et l'Université de Toronto. Ces enquêtes, qui lient techniques documentaires, linguistique et sciences cognitives, mettent en oeuvre d'autres protocoles expérimentaux très élaborés, auxquels nous nous référerons pour émettre quelques propositions en vue d'un nouveau protocole d'enquête, centré sur l'interaction homme/machine et l'étude des stratégies linguistiques mises en place par les utilisateurs d'OPACs.

I- Apports et limites du questionnaire électronique

1- Intérêt des questionnaires électroniques

En élaborant deux questionnaires électroniques⁴⁴, un pré- et un postquestionnaire, le groupe de recherche a suivi une démarche préconisée par la City University de Londres, qui possède une expérience solide dans ce domaine. Micheline Hancock-Beaulieu a montré tout l'intérêt que pouvait présenter une telle démarche pour *l'observation de l'utilisateur*⁴⁵. C'est avec cet objectif d'observation continue que la City University a mis au point le logiciel OLIVE, avec lequel a été réalisée l'enquête de La Villette : ce logiciel possède plusieurs fonctions permettant d'enregistrer la totalité d'une session et d'introduire des questionnaires tout au long de la recherche. Il réalise donc ce que Mme Hancock-Beaulieu appelle une *analyse détaillée [du] comportement [de l'utilisateur] dans le processus complet de la recherche d'informations, et pas seulement pendant sa consultation du catalogue*.

La plupart des enquêtes réalisées aux Etats-Unis ou en Grande-Bretagne sur les catalogues en ligne utilisent les possibilités de l'informatique, comme le rappelait Jean Dickson à la suite d'une étude conduite en 83-84 :

One of the advantages of studying the online catalog is the possibility of collecting data unobtrusively [...] Another advantage of studying the online system is the possibility of using the computer to narrow down or select data.

44- On trouvera une copie de ces deux questionnaires en annexe 3, volume 2, de ce mémoire.

45- HANCOCK-BEAULIEU, M. Les catalogues en ligne jugés par les utilisateurs. *Bulletin des Bibliothèques de France*, 1991, Vol. 36, n°4, p. 314.

L'outil électronique présente deux avantages majeurs :

- celui de supprimer l'enquêteur-médiateur, dont le comportement pourrait influencer l'utilisateur dans ses réponses ; il offre donc certaines garanties d'objectivité durant la collecte des informations.
- celui de fournir des données immédiatement exploitables par le biais de l'informatique, des données homogènes avec celles que peut fournir la session.

Nous pouvons cependant nous demander si les données du questionnaire électronique suffisent toujours dans le cas d'une étude linguistique.

2- Linguistique et questionnaires

2-1 Aspects linguistiques dans les deux questionnaires d'enquête

3 questions ouvertes ont été spécialement intégrées aux questionnaires afin de servir de point de départ à une étude linguistique :

- la question 24 du préquestionnaire
- les questions 8 et 11 du postquestionnaire.

Nous avons volontairement écarté la question 8 de notre étude :

Pour aider quelqu'un à faire la même recherche que vous, sur le système actuel, quels modes de recherche lui conseilleriez-vous de taper ?

Elle présente à nos yeux deux inconvénients :

- l'utilisateur est invité à mélanger dans une même formulation modes et termes de recherche
- 40 de nos 154 sessions possèdent une réponse à cette question, mais 17 seulement présentent à la fois une réponse P08 et une réponse P011, ce qui ne permet pas une étude significative.

La question la plus intéressante, sur le plan linguistique, est incontestablement la question P011 : elle a été conçue pour "libérer" la formulation des contraintes du système et devait permettre d'obtenir une (ré)écriture du sujet en langue naturelle. Nous avons montré au cours de cette étude qu'il n'en a rien été.

2-2 Exploitation de la question P011

Rappelons tout d'abord le taux élevé de non-réponses à cette question: 202 sur 598, soit 34 %. Un faible pourcentage de réponses comporte une énonciation de sujet : 148, soit 25 %.

Le taux de non-réponses peut s'analyser diversement. L'explication la plus rapide consiste à dire que l'utilisateur n'a pas su reformuler sa question : c'est d'ailleurs l'une des hypothèses que se proposait de tester cette question ouverte. Elle serait confirmée par le taux élevé de réponses du type *idem, comme je l'ai fait, de la même façon...* (65 cas) et par le nombre également important de reprises totales ou partielles de formulations antérieures, que nous avons mis en évidence dans le chapitre précédent.

Deux facteurs perturbent cependant cette interprétation.

L'étude a montré que la session exerçait une influence, parfois difficile à évaluer, sur la formulation retenue par l'utilisateur. Mais elle influe également sur sa volonté de répondre. Pourquoi l'utilisateur satisfait du déroulement de la session ou découragé par son résultat chercherait-il une autre formulation de son sujet ou entrerait-il une seconde, voire une troisième fois, le même énoncé ? Certaines réponses sont à cet égard révélatrices :

UG09.4 - *je ne vois pas l'utilité de reformuler ma question*

ME07.16 - *LA FORMULATION N EST PAS REMISE EN QUESTION*

ME14.14 - *ma question est formulée correctement, c'est la réponse qui ne me satisfait pas.*

Cette question, qui semble redondante des questions PR24 et P08, paraît superflue à la majorité des utilisateurs. Plus encore, intervenant après les péripéties de la session, elle attire les commentaires personnels et sert parfois à "se défouler".

Lors de l'élaboration des questionnaires, cette question P011 a fait l'objet de soins particuliers : comment la formuler sans utiliser des concepts que certains utilisateurs risquaient de ne pas comprendre ("langue naturelle") ou sans préjuger de leur réponse ? La formulation retenue ne nous paraît pourtant pas satisfaisante, dans la mesure où elle induit en définitive un double questionnement pour le non-spécialiste : "formulez votre question" et "imaginez un système idéal". Certains commentaires personnels obtenus à la place de la formulation attendue témoignent du caractère subjectif du vocable "système idéal" et des interrogations qu'il suscite (même si quelques-uns relèvent de la plaisanterie). En voici un florilège :

AC17.23 - *en langage naturel et avec une meilleur interface homme-machine*

ME23.2 - *un système idéal est un système personnalisé*

UG16.6 - *EST-CE QUE LE SYSTEME IDEAL EXISTE ?*

UG04.11 - *par la pensée*

ME20.12 - *faites-moi mon devoir pour lundi svp*

Il nous semble donc que cette question, telle qu'elle est rédigée et située dans l'enquête, ne permet pas de tester une formulation libre de l'utilisateur. Inscrite au début de l'enquête (à la place de PR24), elle aurait recueilli davantage de réponses exploitables ; intégrée à un postquestionnaire oral, dans lequel l'enquêteur aurait réorienté la question sur la formulation et non sur le système idéal, elle aurait été encore plus performante. A l'origine, le dispositif prévoyait un questionnaire oral, rempli au terme de la session. Les participants ont effectivement été soumis à ce questionnaire, mais les données, enregistrées de façon manuscrite par différents enquêteurs et non retranscrites, se révèlent, de l'avis même d'un responsable, difficilement exploitables.

2-3 Interprétation des sessions

Dans le mémoire qu'elle consacre au catalogue automatisé de la Médiathèque de La Villette et à l'enquête PARINFO-OPACS, Françoise Belet mentionne que le logiciel Olive permet de *rejouer la recherche devant l'utilisateur, de s'entretenir avec lui, et de recueillir ses commentaires*

pendant cette opération. Elle signale aussi que cette possibilité a été abandonnée car elle aurait entraîné une mobilisation excessive des utilisateurs⁴⁶.

Au terme de notre étude de sessions, nous estimons cependant qu'une étude linguistique complète ne peut s'effectuer sans le regard et les commentaires de l'utilisateur. Ils auraient permis, par exemple, de comprendre pourquoi certains ont choisi d'alterner plusieurs recherches dans la même session ou de multiplier les recherches indépendantes. Ils auraient confirmé ou infirmé les quelques cas d'instanciation qui ont été mis en évidence. L'examen d'une session en compagnie de l'utilisateur aurait surtout servi à justifier chacun de ses choix linguistiques, et peut-être à dégager de véritables stratégies d'interrogation. Le postquestionnaire électronique ne fournit pas le "retour" nécessaire à une interprétation linguistique poussée.

3- Limites du postquestionnaire

Le postquestionnaire a été élaboré dans le but de caractériser l'appréciation portée par l'usager sur sa propre recherche. Il lui est donc demandé d'évaluer sa recherche en 11 questions, auxquelles il ne peut apporter que des réponses très subjectives. Le rapport rédigé sur le projet PARINFO-OPACs signale d'ailleurs cet écueil et remarque en outre que ces questions présentent certaines ambiguïtés⁴⁷:

Le catalogue ne peut être évalué à partir des seules réponses de l'usager ; en effet, la subjectivité de l'utilisateur intervient fortement dans toutes les réponses aux questions concernant l'appréciation portée sur les résultats de la recherche. D'autre part, ce sont ces questions là qui recèlent le plus d'ambiguïté.

Les questions portent en effet sur la satisfaction que l'usager a retirée de sa recherche, les informations qu'il a retenues, son appréciation de la difficulté et des causes de cette difficulté... La confrontation des sessions suivantes permet de mieux apprécier la subjectivité de certaines réponses :

Session UG13.2 :

PR24 = *information sur la technologie des disques optiq*

Recherche en 1 étape et 1 mode de recherche

TIT = *informtions sur la technologie et le*

Réponse : 0

Affichage d'une référence alphabétiquement proche :

"9- L'infrarouge" (2 fois)

Retour à la liste de références proches

Affichage d'une référence : "2- L'infrarouge lointain"

L'utilisatrice, une lycéenne, interroge *juste pour voir comment fonctionne le catalogue.* Alors que les titres balayés ou affichés ne présentent aucun rapport avec la formulation entrée en PR24 ou en première requête, elle se déclare satisfaite de sa recherche, qu'elle juge facile (elle n'a cependant retenu aucun titre).

46- BELET, F., *L'usage du catalogue informatisé : l'exemple de la Médiathèque de La Villette*, p. 36-37.

47- [ECOLE 93], p. 64.

Session UG11.6 :

PR24 = DOCUMENTS SUR LA CONCEPTION ASSISTEE PAR ORDINATEUR

L'utilisateur conduit une recherche en plusieurs étapes et 3 modes de recherche:

- recherche sujet, à partir de CAO

(Il sélectionne plusieurs sujets et affiche 8 références, dont 5 notices complètes).

- recherches titre : *cad report*, *micad* (affichage de la notice), *harvest* (affichage de la notice).

- recherches par mots du titre : (CAO/CAD/conception) ; (conception assistée) ; (conception & assistée)

Au total, 12 références ont été affichées au cours de cette session longue, mais structurée. Alors qu'il a retrouvé la plupart des titres demandés, l'utilisateur, un ingénieur consultant, déclare n'avoir rien trouvé du tout.

On constate, à partir de ces deux exemples, qu'il est impossible de s'appuyer sur les réponses du postquestionnaire pour décider si une recherche donnée a satisfait son auteur. En règle générale, plus un utilisateur est expert de son domaine de recherche, plus sa recherche est ciblée et plus il devient exigeant et capable d'apprécier la pertinence des références trouvées. Mais un utilisateur peut aussi s'estimer satisfait d'une recherche que le professionnel jugera incomplète, le premier ne s'apercevant pas qu'il n'a pas retrouvé certaines références, que le second sait figurer au catalogue.

Seule une comparaison entre le profil et la session de recherche de l'usager d'une part et une recherche type conduite par un ou plusieurs bibliothécaires d'autre part pourrait permettre d'évaluer objectivement une session, et par suite de connaître la pertinence linguistique des choix effectués par cet usager. Ce type d'étude n'était pas envisageable sur un corpus aussi important que celui de l'OPAC de La Villette et n'avait pas été prévu dans le dispositif exploratoire. Une enquête future devrait peut-être envisager de le mettre en place.

II- Quelques réorientations

1- Pour un postquestionnaire oral

Un nouveau dispositif pourrait être étudié, dans lequel le postquestionnaire électronique serait remplacé par un entretien. L'utilisateur - volontaire pour ce type d'enquête - serait invité à relire et expliquer sa démarche à un enquêteur. Celui-ci pourrait disposer d'une grille d'analyse et de codage relativement fine, n'excluant pas cependant de consigner toute remarque utile pour le traitement ultérieur et l'interprétation des données, voire d'enregistrer ces commentaires.

La grille devrait notamment prendre en compte les différentes recherches menées, leur lien ou absence de lien, leur(s) mode(s), leur évolution syntaxique et sémantique ; tenter de mettre en correspondance ce que l'utilisateur a inscrit au clavier et ce qu'il a réellement voulu entrer : a-t-il voulu, entre autres, élargir son sujet, le préciser ? Le terme entré constituait-il pour lui un synonyme d'un terme de recherche précédent ? S'agissait-il d'une expression connue ou a-t-il repris un terme fourni par le système, éventuellement au cours d'une recherche antérieure ?...

Ces questions ne sont que des pistes, qui nous ont été fournies par nos analyses de sessions.

Dans le cadre "dirigé" de ce postquestionnaire, l'enquêteur pourrait alors demander à l'utilisateur de reformuler sa requête, non pas dans un système idéal, mais comme s'il s'adressait à un spécialiste. Cette formulation de la question n'est pas exempte d'une certaine subjectivité, mais ce défaut nous paraît compensé par le fait qu'elle permettrait d'obtenir de véritables énoncés du sujet, et non des remarques personnelles ou des avis sur le système testé. Si, de plus, la même question avait été posée dans un préquestionnaire électronique, il deviendrait possible de comparer, toujours avec l'utilisateur, l'évolution ou la stabilité de sa réponse.

2- Quelle population ?

2-1 Le public de la Médiathèque de La Villette

L'enquête conduite à la Médiathèque de La Villette n'a pas été ciblée sur un public particulier. Le seul "filtrage" mis en place avait pour but d'écartier les enfants, les groupes d'au moins trois personnes et les utilisateurs qui consultaient le catalogue en anglais⁴⁸. Nous avons déjà rappelé qu'il ne s'agissait pas d'un public ordinaire : la Médiathèque offre à la fois des ouvrages de vulgarisation scientifique au grand public et des ouvrages d'histoire des sciences et techniques aux étudiants et aux chercheurs.

48- Quelques principes méthodologiques (extrait du document fourni aux enquêteurs). In [ECOLE 93], Annexe E, p. 1.

Une enquête sur les usagers de la Médiathèque, réalisée en 1991, a fait apparaître 7 catégories de publics :

- les assidus (scolaires et étudiants / actifs)
- les habitués (scolaires et étudiants / actifs)
- les occasionnels (scolaires et étudiants / actifs)
- le "public familles", composé d'enfants accompagnés d'adultes

Dans ces catégories, on peut encore distinguer ceux qui interrogent couramment l'OPAC, que nous avons appelés "experts" du système, et ceux qui l'utilisent peu ; ceux qui connaissent parfaitement le domaine dans lequel ils interrogent et ceux qui le découvrent...

Toutes ces catégories, dont chacune fréquente la Médiathèque et consulte l'OPAC pour des raisons différentes, ont eu la possibilité de participer à l'enquête, sans qu'il soit aisé de les distinguer à la lecture des sessions. Les études du public de la Médiathèque fournissent quelques pistes : les occasionnels sont particulièrement nombreux le week-end ; le grand public fréquente l'étage d'entrée, tandis que la Médiathèque d'Histoire des Sciences se caractérise par un public de chercheurs et de spécialistes⁴⁹.

2-2 Typologie des usagers

Il aurait été intéressant de conduire sur les pré- et postquestionnaires de notre corpus le même type de traitement statistique que celui réalisé à l'ENSSIB⁵⁰. Nous n'avons pu le mettre en oeuvre dans le temps qui nous était imparti pour achever ce travail.

Deux critères nous semblaient particulièrement pertinents pour notre étude linguistique :

- le niveau d'expertise des usagers par rapport au système (que nous avons tenté de prendre en compte en examinant les préquestionnaires)
- le niveau de connaissances de l'utilisateur par rapport au domaine de recherche.

Il s'avère cependant difficile de déterminer si un usager est expert ou non du domaine dans lequel il interroge l'OPAC à partir des questions du préquestionnaire : devons-nous appeler expert celui qui a une formation dans ce domaine, et à quel niveau doit alors se situer cette formation ? Faut-il prendre en compte celui qui travaille ou a travaillé dans ce domaine ? Mais le passionné d'astronomie ou de mécanique automobile n'est-il pas, lui aussi, un spécialiste, même s'il n'a jamais travaillé dans cette branche ? Or, ce type d'expertise n'est pas relevé dans le questionnaire. Nous avons donc renoncé à retenir ce second critère.

Nous pouvons d'ailleurs nous demander si les critères "expertise du système" et "expertise du domaine" ont une réelle influence sur les techniques de recherche et le dialogue homme/machine. Faute d'enquête ciblée sur ces deux critères, on ne peut qu'émettre des hypothèses à ce sujet : la nôtre est que les recherches d'utilisateurs habitués à un OPAC

49- Quelques principes méthodologiques (extrait du document fourni aux enquêteurs). In [ECOLE 93], Annexes E, p. 2.

50- Résultats de l'enquête statistique. In [ECOLE 93], p. 45-74.

donné, sans être nécessairement plus performantes que les autres, sont plus structurées et logiques dans leur évolution. Leurs auteurs font preuve d'une autonomie de langage supérieure par rapport au système. Cette hypothèse reste à vérifier.

Une telle vérification demanderait, elle aussi, la mise en oeuvre d'un second protocole expérimental, dépassant et prolongeant la recherche exploratoire.

2-3 Panel d'utilisateurs

Une nouvelle enquête pourrait être conduite à partir d'un panel d'utilisateurs, panel qui distinguerait au moins les 4 catégories suivantes:

- utilisateurs experts de l'OPAC choisi pour l'enquête et du domaine d'interrogation.
- utilisateurs experts de l'OPAC, mais pas du domaine d'interrogation.
- utilisateurs peu expérimentés ou novices du système, mais experts du domaine
- utilisateurs peu expérimentés du système et novices du domaine.

Il ne nous semble pas nécessaire de prendre en compte l'expertise que les utilisateurs pourraient avoir d'autres catalogues en ligne, du minitel ou de systèmes informatiques, à moins qu'il ne s'agisse d'une expérience acquise lors de la consultation d'un OPAC semblable à celui du test. L'objectif de l'enquête serait en effet d'analyser comment l'utilisateur adapte sa formulation à un système particulier et l'apprentissage qu'il peut faire du langage de ce système.

Elle nécessiterait de fixer, nous l'avons vu, ce que l'on entend par "expert du domaine". Nous pensons qu'une telle enquête serait plus facile à mettre en oeuvre en milieu universitaire, auprès d'utilisateurs étudiant une discipline donnée : il deviendrait possible de distinguer les "novices", étudiants en première année de DEUG⁵¹, et les "experts", étudiants de second et troisième cycles ou professeurs. La notion d'expertise serait ici liée à la formation scolaire et universitaire des utilisateurs.

Ce dispositif pourrait permettre d'examiner la validité des critères "expertise du système" et "expertise du domaine" et de déterminer si ces deux critères ont un réel impact sur l'interactivité qui s'établit entre l'utilisateur et l'OPAC.

L'enquête réalisée à La Médiathèque de La Villette dans le cadre du projet PARINFO-OPACs laisse l'utilisateur libre du sujet sur lequel il interroge. Nos propositions ont jusqu'alors visé à prolonger cette première exploration, qui s'intéresse essentiellement à la formulation "en langage naturel" de l'utilisateur. Bien que mise en oeuvre sur un catalogue en ligne de seconde génération, elle annonce les catalogues de troisième génération et certains systèmes experts de recherche documentaire, qui s'efforcent de traiter des requêtes en langue naturelle et d'apporter à

système OKAPI, développé par le Centre de Recherche sur les systèmes interactifs de la City University de Londres.

En attendant l'exploitation de tels systèmes dans les bibliothèques, une étude linguistique doit aussi fournir une analyse plus fine de l'existant et permettre d'améliorer la convivialité des OPACs en place. C'est pourquoi nous souhaiterions nous intéresser à des dispositifs qui proposent une autre méthodologie pour aborder le dialogue utilisateur / catalogue en ligne.

III- L'évaluation par résolution de problèmes

1- Analyse de quelques dispositifs

De nombreux chercheurs, qui font autant appel aux sciences cognitives qu'à la linguistique, se sont penchés sur les processus de recherche d'information développés par les usagers de catalogues en ligne.

Leurs enquêtes se caractérisent par des protocoles de recherche et une méthodologie différents de ceux qui ont été retenus dans le cadre du projet PARINFO-OPACs : elles ne tentent pas d'analyser une recherche spontanée de l'utilisateur, mais lui proposent au contraire d'utiliser le catalogue pour résoudre des problèmes de recherche d'information imposés. Cette démarche nous paraît intéressante et complémentaire, de par ses objectifs et ses résultats, de celle du projet.

Elle a été adoptée dans de grandes bibliothèques universitaires nord-américaines : notamment par Christine Borgman à l'Université de Californie Los Angeles et de Stanford, par Prudence Dalrymple à l'Université de Wisconsin-Madison et par Joan Cherry et Marshall Clinton de l'Université de Toronto.

1-1 Représentation mentale d'un système de recherche d'information

Christine Borgman, de la Graduate School of Library and Information Science de l'UCLA, University of California, Los Angeles, a conduit une expérience empirique selon cette démarche à l'Université de Stanford, avec pour objectif de tester sa théorie des modèles mentaux ("the mental model theory")⁵².

Son équipe a formé des étudiants non-spécialistes à l'utilisation d'un prototype de système de recherche d'informations basé sur la logique booléenne, qui permet d'interroger une base de données bibliographiques. La formation a eu lieu selon deux méthodes : un modèle conceptuel / des instructions procédurales. Puis on a comparé les représentations mentales que les utilisateurs avaient développées à partir de leur formation au système, au moyen d'une session de recherche sur le catalogue en ligne, complétée par des interviews.

L'enquête consiste en 15 recherches, 5 "simples" et 10 "complexes". (On appelle "recherche simple" une recherche qui nécessite l'usage d'un seul index et d'un opérateur booléen au plus ; une recherche complexe réclame l'usage de deux index ou plus et d'un opérateur booléen au moins.) Il s'agit de recherches bibliographiques, comportant toutes les explications et le vocabulaire nécessaires pour localiser les documents dans une base expérimentale de 6.000 références.

52- BORGMAN, C.L. The user's mental model of an information retrieval system : an experiment on a prototype online catalog. *International Journal of Man-machine Studies*, 1986, Vol. 24, p. 47-64.

Le concept de "modèle mental" est utilisé pour désigner le mécanisme cognitif à l'oeuvre lorsqu'on tente de se représenter un système ou un problème et que l'on infère des connaissances sur ce système ou ce problème.

Cette enquête a fait apparaître des difficultés d'utilisation du système plus importantes que prévu et a permis de mettre en évidence l'efficacité d'une formation par modèle conceptuel pour résoudre des tâches complexes.

1-2 Recherche par reformulation

Prudence Dalrymple a conçu une expérience similaire par ses méthodes à l'Université de Wisconsin-Madison, expérience destinée à étudier les processus de reformulation et le comportement des utilisateurs selon qu'ils consultent un catalogue papier ou un catalogue en ligne⁵³.

Elle a testé 2 groupes de 20 étudiants, chaque groupe consultant l'un ou l'autre des catalogues. L'intérêt de cette expérience réside dans le "retour" qu'elle prévoit sur les données récoltées, au moyen d'un même questionnaire proposé deux fois aux utilisateurs.

Durant la première étape, l'utilisateur est invité à écrire, sans procéder à aucune recherche, de quelle information il a besoin pour résoudre les 6 problèmes posés.

Pendant la seconde, il doit rechercher les informations nécessaires à la résolution de ces mêmes problèmes, soit dans le catalogue papier, soit dans le catalogue en ligne. Il ne dispose pas de ses premières réponses au questionnaire. Ses commentaires durant la recherche doivent être exprimés à haute voix, afin d'être enregistrés. Les réponses écrites sont analysées selon une grille de contenus.

En choisissant ce type de dispositif, de questions "contrôlées", P. Dalrymple estime avoir réduit les risques de variation inhérentes aux recherches fondées sur de véritables questions d'utilisateurs. A ceux qui pourraient juger trop artificielles les questions de ce test, elle répond que, dans un établissement universitaire, la plupart des besoins d'information sont également générés artificiellement, notamment par les cours.

1-3 Enquête expérimentale sur deux types d'instruction pour les usagers de catalogues informatisés

Cette enquête, réalisée en Septembre 90 et rapportée par Joan Cherry et Marshall Clinton, ne porte pas sur un catalogue en ligne - celui de l'Université de Toronto - mais sur une nouvelle formation possible des utilisateurs de ce catalogue⁵⁴ au moyen d'un didacticiel.

Là encore, le dispositif comprend plusieurs groupes d'étudiants, tous issus d'un même cours de première année : un groupe de contrôle, qui n'a reçu aucune formation au système ; un groupe formé au moyen du didacticiel ; un groupe formé de manière classique, durant un cours donné par un bibliothécaire.

53- DALRYMPLE, P.W. Retrieval by reformulation in two library catalogs : toward a cognitive model of searching behavior. *Journal of the American Society for Information Science*, 1990, Vol. 41, n°4, p. 272-281.

54- CHERRY, J.M., CLINTON, M. An experimental investigation of two types of instruction for OPAC users. *The Canadian Journal of Information Science*, 1991, Vol. 16, n°4, p. 2-22.

Les utilisateurs se sont ensuite vus proposer un questionnaire comprenant 9 problèmes liés à leur domaine d'étude (la littérature anglaise).

Exemples⁵⁵:

Question 6 - Vous recherchez des critiques de tragédie. Sous quelle vedette figurent-elles dans le catalogue ?

Tragédie--Critique
 Tragédie--Critiques
 Tragédie--Histoire et critique
 Critiques de tragédie

Question 9 - Combien existe-t-il dans le catalogue de bibliographies sur l'oeuvre de Robert Frost ?

Cette enquête a effectivement permis de montrer l'efficacité du didacticiel et de perfectionner ce type de formation des usagers.

2- Vers un nouveau dispositif ?

Un dispositif comparable pourrait s'avérer très utile pour analyser l'évolution de la formulation durant une recherche, et plus particulièrement pour centrer l'étude sur certaines étapes capitales : par exemple la première requête.

Dans le second chapitre de ce travail, nous nous sommes intéressée à ce passage de la formulation libre du préquestionnaire (PR24) au premier terme de recherche entré par l'utilisateur. Il nous semble, en effet, que cette étape détermine en grande partie le déroulement ultérieur d'une session.

On pourrait, en s'inspirant des exemples précédents, envisager une enquête en trois temps :

- réponse à un préquestionnaire électronique, demandant à l'utilisateur de choisir parmi plusieurs vedettes ou de décrire les informations dont il a besoin pour exécuter une recherche donnée. On pourrait proposer plusieurs niveaux d'interrogation possibles, plus ou moins spécifiques par rapport au sujet.
- consultation du catalogue en ligne pour exécuter une recherche portant sur des sujets connexes.
- réponse à un postquestionnaire oral, permettant à l'utilisateur d'expliquer sa démarche.

Il est probable qu'un tel dispositif permettrait de comparer les choix "spontanés" de l'utilisateur et les termes effectivement entrés au cours de la recherche. Il serait intéressant d'étudier les reprises de termes fournis par le préquestionnaire ou au contraire leur abandon au moment de passer à l'interrogation proprement dite. D'après Joëlle Le

Marec, l'utilisateur d'OPAC établit une telle relation de confiance avec le système qu'il ne cherche pas à optimiser sa recherche en ligne, alors qu'il le fait spontanément quand on l'interroge⁵⁶. Les résultats devraient confirmer ou infirmer l'existence de ce décalage.

L'enquête fournirait une évaluation de la correspondance ou de l'écart linguistique existant entre le langage de l'utilisateur et celui du système. Testé sur le panel d'utilisateurs dont nous avons proposé la constitution (cf. supra 1-5-3), il permettrait de déterminer si les utilisateurs experts du système et/ou du domaine se montrent plus aptes à retrouver spontanément le langage de l'OPAC.

Le projet PARINFO-OPACS a expérimenté en France une méthodologie d'observation du comportement de l'utilisateur qui consulte un catalogue en ligne. Il s'est inspiré des méthodes d'analyse transactionnelle⁵⁷ élaborées par l'un de ses participants, la City University de Londres.

Cette recherche exploratoire a permis de formuler des hypothèses de travail et de dégager des pistes pour de futures expérimentations :

- doubler les questionnaires électroniques d'un dialogue avec l'utilisateur, qui permettra d'avancer dans la "modélisation" des comportements linguistiques durant une recherche d'information.
- mettre en oeuvre une enquête similaire, toujours en milieu "réel", mais sur un panel d'utilisateurs, afin de tester certains critères susceptibles de modifier les caractéristiques linguistiques d'une recherche (tels que le niveau d'expertise par rapport au système et/ou au domaine d'interrogation).
- élaborer une enquête plus ciblée, destinée à mesurer la distance qui existe entre le langage de l'utilisateur et celui du système en certains points clés de la recherche. L'interrogation pourrait alors porter non pas sur une recherche libre, mais sur des tâches de recherche imposées.

Ces prolongements possibles ne répondent pas seulement à des objectifs linguistiques, mais font également appel aux sciences cognitives, aux sciences de l'information et de la communication... Elles sont pluridisciplinaires et s'inscrivent donc bien dans la ligne du projet PARINFO.

56- LE MAREC, J. Les OPACS sont-ils opaques ? La consultation des catalogues informatisés à la BPI du centre Pompidou. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 80.

57- HANCOCK-BEAULIEU, M. Les catalogues en ligne jugés par les utilisateurs. **Bulletin des Bibliothèques de France**, 1991, Vol. 36, n°4, p. 314.

M. Hancock-Beaulieu définit l'analyse transactionnelle comme une méthode d'observation qui consiste à examiner le processus même de l'interrogation, par opposition aux enquêtes d'acceptabilité.

CONCLUSION

Quels sont les enjeux d'une recherche telle que l'enquête PARINFO-OPACs pour les bibliothèques ? Que peuvent en attendre les principaux intéressés, bibliothécaires et usagers ?

A des bibliothèques désormais bien engagées dans le processus d'informatisation, la recherche offre un moyen d'évaluer cet outil de travail que constitue le catalogue en ligne et de mieux comprendre les relations, parfois ambivalentes, qui vont se nouer entre cet outil et leur public.

Le catalogue, quelle que soit sa forme, constitue en effet le pivot d'une bibliothèque, et ce d'autant plus lorsqu'elle n'est pas en accès libre. Il fournit à l'utilisateur *un moyen privilégié d'accès aux collections*⁵⁸, dont il offre une représentation. L'informatisation de ce catalogue introduit des difficultés nouvelles, que le professionnel de l'information perçoit intuitivement, sans toujours avoir les moyens de les évaluer. Une étude "en situation", comme celle de la Médiathèque de la Villette, met ces problèmes à plat et propose des pistes d'étude. Elle offre des points de repère pour l'évaluation d'un OPAC.

Un certain nombre de difficultés dont on rejette aujourd'hui la responsabilité sur l'OPAC freinaient probablement déjà l'utilisateur dans ses consultations du catalogue papier. Mais elles étaient masquées par le fait que le taux de non-réponses n'était pas évaluable et que cet usager n'avait pas le sentiment de recevoir une réponse négative et définitive, ni d'être renvoyé à lui-même. Le catalogue en ligne, parce qu'il amplifie les difficultés de recherche et les blocages, en constitue un puissant révélateur.

Il serait d'ailleurs indiqué de poursuivre les comparaisons des processus de recherche mis en oeuvre par l'utilisateur selon qu'il consulte un catalogue papier ou un OPAC. Rappelons que l'enquête de Prudence Dalrymple, que nous avons brièvement décrite dans le chapitre précédent, fournit quelques pistes intéressantes : elle explique notamment que si les utilisateurs du catalogue en ligne se montrent aussi insatisfaits de leur recherche, c'est parce qu'ils sont beaucoup plus exigeants à son égard, alors que les utilisateurs du catalogue papier gardent une vision concrète de ce qu'il est possible d'obtenir par son intermédiaire⁵⁹.

58- LE LOARER, P. Opacité et transparence des catalogues informatisés pour l'utilisateur. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 65.

59- DALRYMPLE, P.W. Retrieval by reformulation in two library catalogs : toward a cognitive model of searching behavior. **Journal of the American Society for Information Science**, 1990, Vol. 41, n°4, p. 277.

Ces conclusions rejoignent celles de Joëlle Le Marec⁶⁰ ou de Jean Dickson⁶¹, qui observent que l'utilisateur entre dans le catalogue en ligne des termes de recherche qu'il n'utiliserait probablement pas pour rechercher une information dans un catalogue papier : termes trop généraux, recherches au prénom de l'auteur...

Ce constat induit une autre question importante pour le bibliothécaire : faut-il former l'utilisateur à l'OPAC et par quels moyens ? Doit-on lui enseigner les principales procédures, lui fournir un modèle conceptuel du système - au sens que Christine Borgman donne à ce mot - ou bien mettre à sa disposition le référentiel du catalogue ? Les avis divergent sur ce point.

Le véritable enjeu, qui dépasse déjà les OPACs mis en place dans nos bibliothèques, est l'accès au document électronique, un accès qui doit être ouvert à tous. D'où l'intérêt des recherches conduites sur le traitement des requêtes en langue naturelle et les systèmes experts de recherche documentaire.

En attendant la mise au point de ces systèmes, il nous paraît cependant indispensable de réfléchir à des dispositifs de formation des utilisateurs. Les expériences initiées sur le dialogue homme/machine et le comportement de l'utilisateur devraient désormais alimenter cette réflexion.

60- LE MAREC, J. Les OPACs sont-ils opaques ? La consultation des catalogues informatisés à la BPI du centre Pompidou. *Bulletin des Bibliothèques de France*, 1989, Vol. 34, n°1, p. 80.

61- DICKSON, J. An analysis of user errors in searching an online catalog. *Cataloging & classification quarterly*, 1984, Vol. 4, n°3, p. 35.

BIBLIOGRAPHIE

Accès naturel à la base de données DIANEGUIDE. **ECHO : Méthodes d'accès avancées aux services d'information**, Commission of the European Communities, DGXIII, ECHO (European Commission Host Organisation), p. 5-19.

Autre source : [GSI-ERLI 91], p. 81-92

BARTHES, C., GLIZE, P. Planning in an expert system for automated information retrieval systems. In **ACM-SIGIR**, 11th International Conference on Research and Development in Information Retrieval, Grenoble, 13-15 Juin 88. Grenoble : PUG (Presses Universitaires de Grenoble), 1988.

BASSANO, J-C. Un système multi-expert pour la recherche documentaire. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1983, N°9, p. 21-25.

BASSANO, J-C. DIALECT : un système expert pour la recherche documentaire : Thèse Université Paris Sud. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1986, N°23 et 24.
Publication de la thèse d'état, Paris : Université Paris 11, 1986

BASSANO, J-C. Systèmes experts et stratégies de recherche de l'information. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1988, N°32, p. 41-67.

BELET, F. **L'usage du catalogue informatisé : l'exemple de la Médiathèque de La Villette**. Paris : Université de Paris 8, Département de documentation, 1992. 111 f.
Mémoire de DESS "Documentation et technologies avancées"

BERRENDONNER, A. **Cours critique de grammaire générative**. Lyon: PUL (Presses Universitaires de Lyon ; Fribourg (CH) : Editions Universitaires, 1983. 324 p.

BORGMAN, C.L. The user's mental model of an information retrieval system : an experiment on a prototype online catalog. **International Journal of Man-machine Studies**, 1986, Vol. 24, p. 47-64.

BORGMAN, C.L. Why are Online Catalogs hard to use ? Lessons learned from information retrieval studies. **Journal of the American Society for Information Science**, 1986, Vol. 37, n°6, p. 387-400.

BOUCHE, R. Le syntagme nominal, une nouvelle approche des bases de données textuelles. **Meta**, 1989, Vol. 34, N°3, p. 428-434.

BOUCHE, R., LAINE, S., METZGER, J-P. Extraction de connaissances à partir d'une collection de documents. In **Tools of knowledge organization and the human interface**, Congrès organisé par l'ISKO (International Society for Knowledge Organization), Darmstadt (D), 14-17 Août 1990.

CANTER, D., POWELL, J., WISHART, J., RODERICK, C. User navigation in complex database systems. **Behaviour and information technology**, 1986, Vol.5, n°3, p. 249-257.

CARLYLE, A. Matching LCSH and user vocabulary in the library catalog. **Cataloging & Classification Quarterly**, 1989, Vol. 10, n°1-2, p.37-63.

CHAUMIER, J. **L'accès automatisé à l'information**. Paris : Entreprise Moderne d'Édition, 1982. 147 p.

CHAUMIER, J. **Analyse et langages documentaires : le traitement linguistique de l'information documentaire**. Paris : Entreprise Moderne d'Édition, 1988. 186 p.

CHERRY, J.M., CLINTON, M. An experimental investigation of two types of instruction for OPAC users. **The Canadian Journal of Information Science**, 1991, Vol. 16, n°4, p. 2-22.

CHIARAMELLA, Y. Un état de l'art en Informatique Documentaire. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1983, n°12, p. 7-12.

CHIARAMELLA, Y., DEFUDE, B., KERKOUBA, D., BRUANDET, M-F. IOTA : a prototype of an information retrieval system. In **ACM SIGIR Conference**, Pisa (Italy), 1986.

CLEMENCIN, G. Querying the French Yellow Pages : natural access to the directory. **Information processing & management**, 1988, Vol. 24, n°6, p. 633-649.

Autre source : [GSI-ERLI 91], p. 15-31

COURRIER, Y. Analyse et langage documentaires. **Documentaliste**, 1976, Vol. 13, n°5-6, p. 178-189.

CROFT, W.B., THOMPSON, R.H. I3R : A new approach to the design of document retrieval systems. **Journal of the American Society for Information Science**, 1987, Vol. 36, n°6, p. 389-404.

DACHELET, R. Etat de l'art de la recherche en informatique documentaire : la représentation des documents et l'accès à l'information. In INRIA. **Le document électronique, Cours INRIA, 11-15 Juin 1990**. Rocquencourt : INRIA, 1990. p. 107-132.

DALRYMPLE, P.W. Retrieval by reformulation in two library catalogs: toward a cognitive model of searching behavior. **Journal of the American Society for Information Science**, 1990, Vol. 41, n°4, p. 272-281.

DANIELS, P.J. Cognitive models in information retrieval : an evaluative review. **Journal of Documentation**, 1986, Vol. 42, n°4, p. 272-304.

DE BRITO, M. **Réalisation d'un analyseur morpho-syntaxique pour la reconnaissance du syntagme nominal : utilisation des grammaires affixes** : Thèse. Lyon : Université Claude Bernard Lyon 1, 1991. 221 p.

DEBILI, F., FLUHR, C., RADASOA, P. About reformulation in full-text IRS. **Information processing and Management**, 1989, Vol. 25, n°6, p. 647-657.

DEFUDE, B. **Etude et réalisation d'un système intelligent de recherche d'information : le prototype IOTA** : Thèse de doctorat. Grenoble : USMG, 1986.

DICKSON, J. An analysis of user errors in searching an online catalog. **Cataloging & classification quarterly**, 1984, Vol. 4, n°3, p. 19-38.

ECOLE NATIONALE SUPERIEURE DES SCIENCES DE L'INFORMATION ET DES BIBLIOTHEQUES. CENTRE D'ETUDES ET DE RECHERCHES EN SCIENCES DE L'INFORMATION, et al. **Les catalogues en ligne accessibles par le public (OPACs) : recherche exploratoire** : rapport de l'étude réalisée dans le cadre du Programme PARINFO (Programme d'aide à la recherche en information). [Villeurbanne :] [Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques], 1993. Pagination multiple 143-[66] f.

FLUHR, C. Analyse de certaines fonctions que doit remplir un système linguistique dans une utilisation documentaire. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1982, n°5, p. 27-36.

GAUCH, S., SMITH, J.B. An expert system for automatic query reformulation. **Journal of the American Society for Information Science**, 1993, Vol. 44, n°3, p. 124-136.

GSI-ERLI. **Ingenierie linguistique et documentaire : Recueil d'articles**. Charenton-le-Pont : GSI-ERLI, Mai 1991. 166 p.

HANCOCK-BEAULIEU, M. Les catalogues en ligne jugés par les utilisateurs. **Bulletin des Bibliothèques de France**, 1991, Vol. 36, n°4, p. 312-316.

HANCOCK-BEAULIEU, M. Query expansion : advances in research in online catalogues. **Journal of Information Science**, 1992, Vol. 18, n°2. p. 99-103.

IZABEL, Y. **Sur le traitement de l'information en langage naturel : le système SPIRIT**. Mémoire. Paris : Conservatoire National des Arts et Métiers, Institut National des Techniques de la Documentation, 1984. 90 f.

KACPRZYK, J., ZADROZNY, S., ZIOLKOWSKI, A. FQUERY III+ : a "Human-consistent" database querying system based on fuzzy logic with linguistic quantifiers. **Information Systems**, 1989, Vol. 14, n°6, p. 443-453.

KOLMAYER, E. Bases de données grand public et organisation des connaissances. **Bulletin des Bibliothèques de France**, 1992, Vol. 37, n°6, p. 8-13.

LE CROSNIER, H. **Systèmes d'accès à des ressources documentaires : vers des antéserveurs intelligents** : Thèse. Marseille : Université de droit, d'économie et des sciences d'Aix-Marseille III, Faculté des sciences et techniques de Saint-Jérôme, 1990. 355 p.

LE GUERN, M. Sur les relations entre terminologie et lexique. In Actes du colloque Terminologie et industries de la langue. **Meta**, Montréal, Septembre 1989, vol. 34, n°3, p. 340-343.

LE GUERN, M. Un analyseur morpho-syntaxique pour l'indexation automatique. **Le français moderne**, Juin 1991, t. LIX, n°1, p. 22-35.

LE LOARER, P. Opacité et transparence des catalogues informatisés pour l'utilisateur. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 64-77.

LE MAREC, J. Les OPACs sont-ils opaques ? La consultation des catalogues informatisés à la BPI du centre Pompidou. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 78-85.

LJUNGBERG, J. Paraphrasing SQL to natural language. In **RIAO 91** (Recherche d'information Assistée par Ordinateur), **Conference Proceedings Intelligent text and image handling**, Universitat Autònoma de Barcelona, Barcelona, Spain, April 2-5 1991. Centre de Hautes Etudes Internationales d'Informatique Documentaire, 1991, Vol. 2, p. 790-808.

MANIEZ, J. **Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires**. Paris : Les Editions d'Organisation, 1987. 291 p.

MARKEY DRABENSTOTT, K.M., VIZINE-GOETZ, D. Search trees for subject searching in online catalogs. **Library Hi Tech**, 1990, Vol. 8, n°3, p. 7-20.

MEKAUCHE, W., BASSANO, J-C. Analyseur linguistique multi-experts pour la recherche d'informations. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1989, n°36, p. 5-18.

MENON, B. Indexation automatique et intelligence artificielle: quelques questions de stratégie. In Cours INRIA. **Image et intelligence artificielle dans l'information scientifique et technique**. INRIA, 1988, p. 143-175.
Autre source : [GSI-ERLI 91], p. 105-139

METZGER, J-P. **Syntagmes nominaux et information textuelle : reconnaissance automatique et représentation** : Thèse. Lyon : Université Claude Bernard Lyon 1, 1988. 324 p.

MINISTERE DE LA RECHERCHE ET DE LA TECHNOLOGIE, MINISTERE DE L'EDUCATION NATIONALE, DE LA JEUNESSE ET DES SPORTS. Programme d'aide à la recherche en information (PARINFO). **Brises**, 1991, n°16, p.81-82.

MITEV, N., HILDRETH, C. Les catalogues interactifs en Grande-Bretagne et aux Etats-Unis : Systèmes et interfaces. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 22-33.

NIE, J. **Un modèle logique général pour les Systèmes de Recherche d'Informations: application au prototype RIME** : Thèse. Grenoble : Université Joseph Fourier Grenoble 1, Laboratoire de Génie Informatique IMAG, 1984. 238 p.

NORMIER, B. Interfaces en langage naturel avec des bases de données. **Télécom**, n°87, 1991, p. 36-39.
Autre source : [GSI-ERLI 91], p. 7-10

PUJO, P. **Développement d'une interface conviviale pour l'interrogation en langage naturel d'une base de données avec utilisation des concepts et des moyens de l'intelligence artificielle** : Thèse. Paris : Université Paris XI, 1989.

RADASOA, H.P. **Méthode d'amélioration de la pertinence des réponses dans un système de bases de données textuelles** : Thèse. Paris : Université de Paris Sud (Orsay), 1988. 156 p.

Répertoire des services Télétel MGS. **La lettre de Télétel**, n°18, 3ème trimestre 1989, p. 12-15.
Autre source : [GSI-ERLI 91], p. 11-14

ROLLAND, C., PROIX, C. Une approche linguistique pour la conception des systèmes d'information. In **Les systèmes d'informations élaborées. Congrès, Ile Rousse, 1989-05-31**. Paris : Université Paris 1, Société C.R.I.L., 1989, p. 17-40.

ROUAULT, J. Linguistic methods in information retrieval systems. In **Advances in intelligent retrieval - Informatics 8**. London: ASLIB, 1985.

SARACEVIC, T., KANTOR, P. A study of information seeking and retrieving. II, Users, questions and effectiveness. **Journal of the American Society for Information Science**, 1988, Vol. 39, n°3, p. 177-196.

SINNO-RONY, S. Les hypercatalogues : nouvelles perspectives pour les OPACs. **Bulletin des Bibliothèques de France**, 1991, Vol. 36, n°4, p. 303-311.

VAN SLYPE, G. **Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires**. Paris : Les Editions d'Organisation, 1987. 277 p.

VICKERY, B.C. Knowledge representation : a brief review. **Journal of Documentation**, 1986, Vol. 42, n°3, p. 145-159.

VICKERY, A., BROOKS, H., ROBINSON, B., VICKERY, B. A reference and referral system using expert system techniques. **Journal of Documentation**, 1987, Vol. 43, n°1, p. 1-23.

VIZINE-GOETZ, D., MARKEY DRABENSTOTT, K.M. Computer and manual analysis of subject terms entered by online catalog users. In **ASIS, Annual meeting, 54, Washington DC USA, 1991-10-27, Systems understanding people**. Washington, 1991, p. 156-161.

WITT, M. Lisibilité de l'OPAC. **International Cataloguing & Bibliographic Control**, 1989, Vol. 18, n°4, p. 51-56.

WITT, M. Simplification ou complication ? Quelques réflexions sur la constitution d'un catalogue en ligne à la Médiathèque de La Villette. **Bulletin des Bibliothèques de France**, 1989, Vol. 34, n°1, p. 48-59.

ZARRI, G.P. Etat de l'art : les nouvelles tendances de l'informatique documentaire. **Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire**, 1988, n°32, p.11-40.

BIBLIOTHEQUE DE L'ENSSIB



9653653