

**Denise MALRIEU**

## **Les apports d'une étude différentielle de la demande bibliographique pour la modélisation des utilisateurs**

Une longue tradition nous a accoutumés à réduire le signifié au concept, et à estimer que la "chose à dire" pourrait jouir de quelque autonomie à l'égard de la manière dont elle est dite. Dès que l'on a affaire à un corpus, cette illusion devrait se dissiper d'elle-même.

F. Rastier, *Sens et textualité*

### ***I. Introduction***

Les difficultés de communication utilisateurs/système dans l'interrogation des banques de données sont à l'origine de nombreux travaux de développement en informatique et IA : mise au point de menus guidant l'interaction, d'interfaces en langage naturel, introduction des modèles de la similitude, assistance à la consultation de thésaurus (Shoval 85, 86), modélisation de l'utilisateur par informations externes à la tâche (Rich, 79 ; Chiaramella, Defude, 87 ; Pejtersen, Austin, 83) ou par diagnostic à partir de ses erreurs (McCoy, 89).

Ces travaux dans leur écrasante majorité prennent comme point de référence le système et non les usagers pour lesquels il est censé fonctionner. Ils montrent les "erreurs" fréquentes des usagers par rapport

aux stratégies optimales prévues par le système ; l'utilisateur est défini par son inadaptation à ce dernier, et une énergie impressionnante est dépensée pour faciliter son adaptation par la construction chez lui, du modèle mental exact du système. Or ces systèmes sont le plus souvent des constructions a priori, héritières d'outils antérieurs (ici les outils bibliothéconomiques), d'une psychologie naïve de "l'Usager", et soumises à des contraintes de coût technologique.

Si l'approche cognitive de la modélisation de l'utilisateur s'est développée à propos de domaines de connaissances restreints, dans l'interaction avec des bases de données factuelles gérant un savoir "vrai" (Pollitt, 87, Paris, 88), ce n'est pas le cas dans le domaine des banques de données bibliographiques scientifiques (les BDB) qui constituent notre objet d'étude : celles-ci gèrent la représentation de produits textuels renvoyant à une multiplicité de constructions des savoirs, liées à des champs et pratiques scientifiques hétérogènes.

Les approches évaluatives des BDB dans les années 70 (voir par exemple Martin et al. 73), représentaient un effort pour basculer le point de référence du côté de l'utilisateur ; elles en sont restées à des constats quantitatifs (généralement sévères), et se sont révélées inaptes à une définition opératoire de la "pertinence" de l'information fournie en sortie. Ceci pour deux raisons : i) en l'absence d'une modélisation adéquate des usagers, les critères de pertinence paraissaient trop variables d'un utilisateur ou d'un champ à l'autre ; ii) il manquait une approche cognitive précise des processus d'interprétation textuelle aux deux étapes auxquelles elle s'effectue : dans le traitement de condensation documentaire (résumé et indexation et théorie implicite du texte à laquelle ces processus renvoient) ; et lors de la lecture par les usagers de cette représentation condensée (inférences effectuées par ces derniers selon leur niveau de formation).

Une approche ergonomique de la *conception et consultation* des BDB scientifiques ne peut se limiter à la simple observation de l'interaction de

l'utilisateur avec le système existant et à l'élaboration d'interfaces sophistiquées entre des usagers que l'on connaît généralement mal et un système<sup>1</sup> dont les choix épistémologiques ne sont pas réellement explicites. Elle implique au préalable une démarche de projet, c'est à dire une analyse des situations et comportements réels de recherche d'information et de lecture scientifique dans leurs différents aspects : routines d'exploration textuelle en fonction des genres textuels et des objectifs de lecture, opérations de sélection, surlignage, mises en relations, extraction de mots clés, etc.).

Les travaux de Belkin et al.<sup>2</sup>, entre autres, ont dégagé de l'observation de la négociation documentaliste/utilisateur lors de l'interrogation, les différents éléments de la construction du modèle de l'utilisateur par le documentaliste.

*Objet de l'analyse :*

Nous supposons ici qu'avant toute interaction de négociation, le travail du documentaliste est orienté par la façon dont l'utilisateur formule sa demande. Nous avons donc choisi d'analyser sa formulation première et spontanée avant toute négociation. Notre analyse a porté sur un corpus de 150 demandes adressées par correspondance sur papier libre à un centre de banques de données en sciences sociales du CNRS. Ce corpus permet de mettre en évidence l'ensemble des éléments d'information que l'utilisateur juge nécessaire de fournir à son interlocuteur pour guider son interprétation de la demande et d'en inférer les modèles implicites de la

---

<sup>1</sup> On entend ici par "système" *l'ensemble des composantes* qui engendrent la BDB : institution documentaire, analystes-indexeurs qui assurent le traitement intellectuel, système informatique.

<sup>2</sup> Belkin a découpé les énoncés du dialogue usager-documentaliste en focus qui instancient les 23 sous-buts regroupés en catégories dont : détermination des objectifs de l'utilisateur ; de son statut ; de son état de connaissance du champ ; de son expérience du système, de ses activités bibliographiques antérieures ; de la description du problème ; des types de documents souhaités, etc.

demande bibliographique mis en jeu spontanément par les différents usagers non guidés par l'intermédiaire humain. Elle permet aussi d'évaluer la distance entre le modèle spontané de la demande et les exigences du système interrogé.

De plus, la pratique de l'interrogation nous ayant fait toucher du doigt une hétérogénéité des publics dans leur capacité à formuler les demandes adéquates pour le système, il paraissait intéressant de caractériser de façon plus précise cette distance au système. Dans l'optique d'une *approche différentielle de la demande*, contrastant avec l'hypothèse d'un modèle général unique qui prévaut dans la conception du système analysé, nous avons défini et comparé 5 groupes d'usagers selon leur statut professionnel et/ou leur niveau de formation : licence (19), maîtrise (27), DEA et thèse (32), enseignants-chercheurs (EC : 33), professionnels (autres que EC : 37) et analysé les demandes selon cette partition.

*La démarche suivie :*

Nous sommes partie du projet suivant : dans l'optique du développement du dialogue Homme/Machine en langage naturel et d'interfaces intelligents prenant en compte la modélisation de l'utilisateur, nous faisons le postulat qu'il est possible d'identifier des caractéristiques du fonctionnement cognitif de la demande à travers une analyse de la formulation discursive. Peut-on définir les différents outils de linguistique textuelle (ou infratextuelle) qui permettent de repérer dans le texte de la demande les éléments nécessaires à la modélisation de l'utilisateur ?<sup>3</sup>

---

<sup>3</sup> Nous renvoyons à Malrieu (à paraître) qui concerne les apports de la linguistique textuelle et de la pragmatique à la modélisation du statut et de la distance sociale de l'émetteur au service sollicité (mode d'établissement du contrat, de sollicitation, variation des schémas énonciatifs de la demande, qui n'est pas sans influencer sur la place de l'énonciation thématique dans la structure discursive (influence du dialogique sur le tactique (linéarité du discours) et sur le thématique). Nous n'aborderons pas non plus la façon dont l'analyse textuelle permet de définir à travers une dialogique interne à l'émetteur, des indicateurs sur le statut épistémique du demandeur : les types de vouloir savoir hétéronome

*Les types d'analyses :*

Nous centrant dans cet article sur l'analyse de la *thématique* de la demande, notre objectif est de caractériser les distances entre les formulations spontanées des demandes documentaires de différents groupes et le mode de gestion des données par le système informatique sollicité, essentiellement le repérage : i) des éléments constitutifs de la thématique ; ii) du niveau de spécialisation de la demande à travers le lexique utilisé et de la façon dont ce dernier est éventuellement reformulé pour la recherche bibliographique (pour en inférer des éléments sur la représentation du système par l'utilisateur (modèle mental du système)) ; iii) étude des distances, différentes selon les groupes, au lexique du système et des transformations de ces demandes par le documentaliste pour les rendre acceptables par le système. Cette caractérisation du positionnement des différents groupes par rapport au système peut par la suite aider à la mise en place d'une assistance différentielle à l'interrogation.

***II. Les éléments constitutifs du thème de la demande et de la classification documentaire***

Les banques de données résultent d'une activité multiple de classification : un document est à la fois un objet physique (ouvrage, revue, thèse), un genre d'élaboration, le compte-rendu d'une activité scientifique ; il est affecté à un plan de classement préétabli qui définit des domaines (disciplines) et spécialités, qui se spécifient dans les mots clés en objet d'étude. Les compte-rendus de l'activité scientifique traités dans le système documentaire relèvent d'une description en schémas

---

ou autonome (Coquet, 1984), rapports du sujet à son objet d'étude, types de questionnements à l'origine de la demande.

D'autre part une analyse a été faite des activités de paraphrasage comme des types d'ambiguïtés linguistiques fréquents dans ce discours de demandes bibliographiques.

prototypiques pour chaque discipline (théorie, objet d'étude, technique et méthode, culture en préhistoire ; objet d'étude, théorie, méthode et procédure, population étudiée en psychologie). La recherche d'information dans les BDB suppose une plus ou moins grande maîtrise de ces variables du schéma de description du document. Il est donc intéressant d'analyser dans quelle mesure celles-ci sont présentes dans la formulation de la demande et dans le système (voir en Annexe 1 les contraintes du système).

Nous avons analysé les critères explicités par l'utilisateur pour sélectionner les documents qu'il recherche (voir en Annexe 2 deux exemples de requêtes). Ces critères peuvent correspondre à des dimensions de description des documents, thématiques ou non thématiques. Le nombre total de critères différents observés sur l'ensemble des lettres s'élève à 34 dont 26 thématiques et 8 non thématiques. Nous appellerons *clés de tri* les critères opérant dans le système.

### *2.1. Les critères non thématiques*

Ils concernent des caractéristiques non liées au contenu (dates, langue des documents) ou portant sur une typologie des textes (genres d'élaboration, de publications, de matériel source utilisé), type de recherche (étude de cas, recherche expérimentale etc.).

Les critères non thématiques sont de façon générale *sous-utilisés* par rapport à ce que le système exige de préciser obligatoirement. Aucune de ces clés n'est utilisée par le quart des utilisateurs : les trois clés les plus présentes (utilisées par 12 à 16% des demandeurs) concernent par ordre croissant les dates de publication des documents (la fourchette chronologique à explorer), les types de documents (ouvrages, thèses, articles) et les langues des documents. D'autre part, malgré la faiblesse de leur usage, la palette des critères non thématiques produits par l'ensemble des usagers est *plus riche* que ce que le système propose. On touche ici à la différenciation des genres qui est plus précise chez les usagers que dans le système. Ceci a des implications sur la capacité du système à adapter sa réponse à la diversité des usagers (en particulier par les types de publications et leurs niveaux d'élaboration).

<b>Clés Non Thém.</b>	Prof.. (37)	Ens/Cher (33)	DEA (32)	Maîtrise (27)	Licence (19)	Total (148)	% (148)
=Nb max réf.	1	1	-	-	-	2	1,3%
=Langues doc	7	9	5	3	-	24	16%
=Dates doc.	8	10	-	-	1	19	13%
Type de doc.	7	3	3	4	5	22	15%
+Type de publicat.	1	3	3	1	2	10	6,7%
+Type de recherche	-	2	-	-	-	2	1,3%
+Matériel source	-	-	-	1	-	1	0,6%
Zone de tri	1	1	1	-	-	3	2%
Localisation doc	-	1	-	1	-	2	1,3%
<b>TOTAL</b>	<b>25</b>	<b>29</b>	<b>12</b>	<b>10</b>	<b>8</b>	<b>83</b>	
<b>Clés thématiques</b>							
=Nom BDB	4	7	1	-	-	12	8%
Discipline	7	15	13	15	4	54	36,4%
Spécialité	11	7	11	8	6	43	29%
=Descript. thémat.	23	28	26	21	17	115	77,7%
=Aires géo.	10	13	12	10	5	50	34%
Période historique	4	7	5	4	-	20	13%
Nom de culture	-	-	-	1	-	1	0,6%
Ethnie	-	3	2	2	-	7	4,7%
Pop. CSP	-	-	2	1	3	6	4%
Pop. Age	-	3	2	3	4	12	8%
Pop. Inadapt.	3	-	2	3	3	11	7,4%
+Cadre écologique	3	1	1	5	5	15	10%
Nom d'institution	1	1	-	1	-	3	2%
Niveau d'enseign.	2	6	3	1	-	12	8%
Genre littéraire	-	-	4	-	-	4	2,7%
Nom d'auteur étudié	4	4	4	6	-	18	12%
Nom pers. hist.	1	-	1	1	-	3	2%
Nom oeuvre étudiée	-	2	1	-	-	3	2%
+Courant/Ecole	-	-	1	2	-	3	2%
Théorie	-	-	1	1	-	2	1,3%
Technique mesure	-	1	1	-	-	2	1,3%
+Type stimuli	-	2	-	-	-	2	1,3%
+Niveaux explication	-	1	-	-	-	1	0,6%
Technique therap.	2	-	-	1	-	3	2%
Techn. évaluatives	-	1	-	-	-	1	0,6%
Nom de revue	2	1	-	-	-	3	2%
Nom langue étudiée	1	1	2	-	-	4	2,7%
Nb clés non thémat. différentes	6	7	4	4	3	8	
Nb clés thémat. différentes	15	18	20	18	8	26	

= : clés obligatoires pour le système

+ : clés produites par les usagers inexistantes dans le système

*Tableau 1. Usage des clés de tri selon les groupes*

Enfin, l'usage des critères correspondant aux clés de tri du système semble lié à la pratique de la recherche documentaire, puisqu'elles sont essentiellement présentes chez les "professionnels" et enseignants-chercheurs, alors que les groupes en formation (y compris DEA et thèses) les ignorent.

## *2.2. Caractérisation des clés de tri thématiques*

A l'intérieur des critères liés à la thématique de la demande, nous distinguerons trois catégories auxquelles nous assignons des fonctions cognitives différenciées :

- Le lexique lié à *l'approche disciplinaire* regroupe les noms de discipline, de spécialités, les noms de BDB dans la mesure où ceux-ci véhiculent généralement une information disciplinaire. La définition préalable du point de vue sur l'objet (*Protée*, 1988), et du *domaine de connaissance* à explorer constitue une partie obligatoire de la démarche d'interrogation et aura pour effet de réduire les polysémies par la définition d'un domaine (Rastier 87).

Dans l'échantillon analysé, on constate que 38% des sujets nomment une *discipline*, 25,6% une spécialité ; 64,8% spécifient l'approche privilégiée par un nom de discipline ou de spécialité ou de BDB.

- Les termes désignant le thème de la bibliographie ou l'objet d'étude, que nous désignerons comme "*descripteurs*" sans qu'il y ait référence au lexique du système.

- Les clés de tri contextuelles spécifient des déterminations sur l'objet (population, pays, ethnie, période, nom de langue étudiée etc.) et font partie d'un lexique d'encyclopédie de base.

Alors que le système ne différencie pas les dimensions de description de la question, nous avons choisi de décomposer la formulation du thème de recherche ou de la bibliographie en catégories sémantiques qui désignent par un *superordonné les facettes de description de la question*. Pour l'ensemble des requêtes nous avons distingué 26 catégories qui relèvent de domaines sémantiques différents et ont des fonctions cognitives différentes : à côté des "descripteurs thématiques" qui désignent le corps notionnel de l'objet d'étude (cf. l'étude des descripteurs), nous avons distingué les catégories lexicales qui apportent des déterminations complémentaires. L'usage de ces catégories et leur statut (comme thème central ou comme détermination du thème central) est évidemment lié à la discipline. La distinction entre descripteurs et clés de tri contextuelles doit permettre une première évaluation du degré de conceptualisation de la question : l'effort de lexicalisation de l'objet d'étude à travers les descripteurs est un indice de l'effort de définition et d'explicitation des concepts centraux pour décrire le problème au destinataire, fonction que n'ont pas les clés de tri contextuelles. On peut considérer que les descripteurs ont une fonction de définition des classes en intension et que les clés contextuelles spécifient des



sous-classes par ajout non de traits plus spécifiques mais de classes d'application qui relèvent du langage courant, d'un lexique de non spécialiste (ce qui ne signifie pas que le choix de ces classes d'application n'ait pas des incidences sur la définition en intension du signifié des descripteurs, mais ces incidences ne sont pas gérées par le système).

### 2.3. Les différences selon les groupes d'utilisateurs

La longueur (en nombre de mots) de la formulation de la thématique croît avec le niveau de formation, si l'on regroupe discipline, thème de recherche, thème de bibliographie.

Le nombre de clés de tri thématiques différentes produites dans chaque groupe n'est pas homogène : il est à peu près le même chez les DEA, maîtrise et enseignants-chercheurs (18 à 20) ; il est plus faible dans le groupe des "professionnels" (15) et chute à 8 en licence. La richesse de la description thématique est bien liée au niveau de formation universitaire comme le confirme la répartition du nombre de clés de tri de chaque demande (de 1 à 5) selon les groupes.

	Professionnels	Licence	Maîtrise	DEA	Enseignants-Chercheurs
Long. Moyenne	25,5	26,4	40,7	49	56,2
Clés non thémat.	0,67	0,44	0,33	0,38	0,87
Clés thémat.	2,10	2,61	3,18	3,06	3,15
Nb total de clés	2,78	3,05	3,51	3,45	4,03
Voc. Discipl.(% Ss)	46	50	81,4	59,3	81,8

Tableau 2. Longueur moyenne de la thématique en nombre de mots et Nombre moyen de clés de tri non thématiques et thématiques

Le calcul du X<sup>2</sup> pour les clés thématiques donne une valeur de 45.59 > 32 à p= .01. La variable discipline n'influe pas sur le nombre de clés de tri thématiques utilisées. (X<sup>2</sup>= 26.13, NS.30 < p < .50).

Précisons que ces chiffres portent sur la présence ou l'absence de chacune des clés de tri pour une même demande et non sur la richesse du lexique utilisé dans chacune des clés.

On peut en tout cas constater d'une façon globale le *faible nombre de clés de tri combinées* dans la formulation des demandes, puisque pour l'ensemble de l'échantillon 1 à 2 clés thématiques représentent 41,4%, 3 clés 34%, 4 clés 15% et 5 clés 9,5%.

Si l'on compare les groupes quant à leur usage d'un lexique d'orientation *disciplinaire*, on constate qu'il est le plus faible (46% des cas) chez les "professionnels" qui, dans 40% des cas, définissent leur question par des descripteurs ou des clés de tri génériques autres que des descripteurs (29% dans le présent échantillon). Ce public non universitaire ferait donc moins facilement référence au découpage académique du savoir. L'usage du vocabulaire disciplinaire croît des professionnels à la licence, au DEA, à la maîtrise et aux enseignants-chercheurs (voir tableau ci-dessus).

La distinction que nous avons introduite entre descripteurs et clés de tri contextuelles permet une première évaluation du degré de conceptualisation de la question. Dans 21,5% des cas les usagers posent leur question sans descripteurs, donc sans aucune précision notionnelle. La répartition de ce type de questionnement selon les groupes d'usagers montre une tendance à varier avec le niveau de formation ( $X^2 = 9.2$  ;  $.05 < p < .10$ ).

#### *2.4. Usage comparé des clés de tri par les utilisateurs et par le système*

Ne pouvant accéder que de façon partielle aux recherches bibliographiques éventuellement effectuées par le service documentaire sur l'échantillon de requêtes observé, nous avons pris au hasard 35 requêtes bibliographiques traitées par ce dernier, dans le domaine des sciences de l'éducation et de la psychologie de l'éducation.

Nous avons effectué le même relevé des clés de tri utilisées que pour notre échantillon de demandes. En ce qui concerne l'usage des clés de tri non thématiques, le système précise toujours les dates à interroger et le plus souvent les langues des documents à fournir ; pour les clés de tri thématiques le système précise toujours la ou les BDB à interroger. De même, il existe toujours des descripteurs.

Le nombre de clés de tri utilisées est en moyenne de 3,57 alors qu'il était de 2.69 pour les requêtes de l'échantillon dans les mêmes domaines, tous groupes confondus.

Le nombre de clés de tri différentes s'élève à 15 pour le système, alors qu'il était de 18 en psychologie et de 13 en sciences de l'éducation dans l'échantillon des demandes.

Si nous comparons la répartition des nombres de clés de tri thématiques pour les différents groupes et dans le système, nous constatons que la différence entre système et groupes est significative pour les professionnels ( $X^2 = 19.34$ ,  $p < .01$ ), et licence, mais non significative pour les autres groupes (DEA :  $X^2 = 5.36$ ,  $.20 < p < .30$  ; maîtrise :  $X^2 = 6.74$ ,  $.10 < p < .20$  ; enseignants. chercheurs :  $X^2 = 8.84$ ,  $.05 < p < .10$ ).

On peut utiliser ces données comme indice du fait que la distance au système soit plus grande pour les populations moins proches de la norme universitaire. Ceci a des implications quant aux facilités d'accès au système pour les publics tout venant (comme semblent le nier les publicités pour l'interrogation sur Minitel).

#### *2.5. Conclusions concernant l'usage des clés de tri thématiques*

i) - Une bonne minorité d'usagers (35%) issue surtout de groupes non universitaires ne pense pas à préciser l'approche disciplinaire, qui est obligatoire pour interroger le système de façon efficace.

ii) - Plus du cinquième de l'échantillon ne définit pas de descripteurs.

iii) - En ce qui concerne la combinatoire des clés thématiques, le modèle le plus fréquent consiste en : Discipline ET Descripteurs ET Déterminations d'ordre spatio-temporel ou de type de population. On peut faire l'hypothèse qu'il s'agit là d'un *schéma cognitif de base*, minimal, de la demande documentaire qui se retrouve dans la structure des titres nominalisés des sujets de recherche. Il correspond dans sa forme la plus pauvre au schéma de la demande documentaire en bibliothèque ; ex. : "je cherche des documents sur la peinture orientaliste du XIX<sup>e</sup> siècle".

iv) - La richesse de la désignation de l'objet, évaluée par le nombre de critères différents utilisés croît avec le niveau de formation, certains critères utilisés par les chercheurs n'étant d'ailleurs pas présents dans l'indexation du système (niveaux d'explication, types de stimuli). Nous allons maintenant examiner quelques indicateurs linguistiques du fonctionnement cognitif de la demande selon les niveaux de formation.

### ***III. Evaluation des niveaux de généralité de la formulation***

#### *3.1. Analyse du lexique des requêtes et niveau de généralité : la notion de niveau de base*

Les études de Chen et Dhar (1990) par exemple, ont montré que les groupes en formation ont tendance à adresser leur requête au système en mobilisant le modèle de la demande en bibliothèque, en choisissant par prudence une formulation très générale avec un schème opératoire d'économie : le plus général appellera le plus spécifique. Il paraît intéressant d'explorer si le niveau de généralité varie avec les groupes et quelle est sa distance au niveau choisi par le système.

Dans sa discussion du niveau de base, Rosch précise (1978 : 43, Mervis et Rosch, 1981) que le niveau d'abstraction auquel un objet sera

considéré varie selon le contexte et selon l'expertise<sup>4</sup> (pour une discussion du niveau de base et expertise voir D. Dubois, 1991 ; Tanaka et Taylor, 1991). On peut ajouter que l'activité de catégorisation et de dénomination des catégories est toujours dépendante des nécessités de différenciation en fonction du type de tâche, des objectifs poursuivis et du type d'interlocuteur dans la communication (voir Cruse 1977) ; peut-on définir un niveau de base dans le système, par rapport auquel les populations d'utilisateurs se situeraient à des distances variables ?

Nous pouvons considérer que le schéma de base décrit plus haut ne sera pas spécifié de façon identique par les différents groupes et par le système. On fera l'hypothèse que, contrairement aux clés de tri contextuelles qui apportent des spécifications spatio-temporelles et font partie d'un savoir stable et partagé, les descripteurs qui lexicalisent l'objet d'étude dans son aspect notionnel seront sensibles aux différences de niveaux de formation.

*L'étude du niveau de spécificité du lexique* peut être abordée à travers l'analyse des syntagmes nominaux. Dans ses études sur le niveau de base, E. Rosch a mis en évidence les rapports entre lexicalisation et niveau du concept dans les taxinomies : au niveau de base correspondrait une lexie simple (table, chaise, chien), alors que les niveaux plus spécifiques se lexicaliseraient en lexies complexes (table de nuit, chaise à porteur, chien de chasse). A travers *l'étude formelle* des structures syntaxiques du lexique on testera l'hypothèse d'une complexification de la structure des syntagmes nominaux (SN) avec le niveau de formation, liée au processus de spécialisation de la terminologie et de nominalisation

---

<sup>4</sup> "The principles underlying the determination of which hierarchical level is basic are expected to be universal. However, for a given domain, the particular level which is found to be basic may not be universal. This level can vary as a function of both the cultural significance of the domain and the level of expertise of the individual... Dougherty and Rosch et al both provide examples of the relativity of the basic level." in *Categorization of Natural Objects*, 89-115.

propre au langage scientifique : davantage de phrases nominales, augmentation des enchâssements de SN dans l'énonciation de la requête.

### 3.1.1. Etude de la structure des SN

Pour des raisons de complexité de la tâche, on a sélectionné dans le corpus des 150 demandes, trois groupes de 10 sujets chacun (enseignants-chercheurs, professionnels, licence), dont le thème relève de la psychologie de l'éducation.

Préalablement, on notera que l'énonciation de la requête se fait à l'intérieur de *structures discursives* différenciées :

- les étudiants de licence et les professionnels ont tendance à exprimer leur requête à l'intérieur de la phrase de demande ou de présentation du thème de recherche (ex. : "je travaille actuellement sur les enfants dyslexiques du milieu algérois") ;

- les enseignants-chercheurs énoncent plus souvent leur requête par la présentation graphiquement structurée d'une liste de points énoncés sous la forme de phrases nominales ou de mots clés. Chez ces derniers le nombre de phrases nominales est de 23, contre 10 le nombre de phrases verbales, alors que les deux s'équilibrent chez les professionnels (12 et 13).

Si l'on considère *la complexité de la structure syntaxique des SN*, on constate que le *nombre de syntagmes nominaux simples* (Nom seul ou Nom +Adj) est plus important en licence que chez les enseignants-chercheurs ou les professionnels. Si l'on observe les *SN complexes* (composés de 2 SN ou plus : génitif ou syntagme prépositionnel) pouvant exprimer des formes de détermination plus complexes que la forme adjectivale, on constate une décroissance du groupe enseignants-chercheurs. au groupe professionnel, et à la licence. Il est notable que cette différence est liée à un emploi plus fréquent dans le premier groupe des formes déverbales (perception de l'élève par l'enseignant) (cf. le décompte du lexique exprimant les processus psychologiques qui prend le plus souvent la forme de Nom avec rection verbale (31 chez les enseignants-chercheurs contre 20 chez les professionnels et 25 en licence).

Si nous comparons le degré de complexité des phrases nominales (ie le nombre de syntagmes nominaux les composant) nous constatons là aussi une différence selon les groupes :

	SN simples	SN Compl .	2 SN	3 SN	4 SN	5 6-8 SN	SN
Ens. Cherch.	15	58	34	6	8	1	3
Professionnels	12	45	22	4	5	4	2
Licence	32	39	18	10	2	1	1

Tableau 3. Complexité syntaxique des descripteurs

Ces données sont confortées par l'analyse de la structure syntaxique des "mots clés". Nous analysons ici dans le corpus des 150 demandes le lexique des "mots clés" désignés explicitement comme tels par les sujets.



### 3.1.2. La syntaxe interne aux mots clés

Existe-t-il un modèle du mot-clé, dont le parangon serait le nom propre (qui interdirait l'usage de l'article et de prépositions de détermination), ou bien observe-t-on une grande variabilité des formes syntaxiques ?

a) *L'usage des déterminants et de l'article défini* : les syntagmes nominaux sont massivement produits sans article (bilinguisme, anorexie) ; quelques cas (de singulier obligatoire) le réintroduisent : "la question agraire en France", ou dans les syntagmes prépositionnels : "transfert dans les apprentissages", et dans les génitifs : "histoire des ports", ou le déverbal : "apprentissage de la lecture".

b) *La préposition "de" (génitif)* : elle est massivement conservée ; on observe seulement quatre cas de suppression après déverbal ("restructuration emploi"), ou de détermination adjectivale ("sociabilité enfants"), ce qui confirme l'existence chez les usagers du modèle du mot-clé sans syntaxe, modèle qui était dominant dans les débuts des BDB.

c) *Les prépositions locatives* : "dans, à, en, chez" : on trouve 14 cas (contre 15) où la préposition de lieu est supprimée. On voit là que les sujets utilisent des structures de détermination complexes qui sont interdites dans le système (où la préposition locative est remplacée par le ET booléen et correspond souvent à une clé de tri contextuelle).

d) *La complexité des SN* : les formes les plus fréquentes sont : Nm, Nm Adj, Nm de Nm, mais il existe aussi des SN plus complexes (Nm déverbal +Prép+Nm Adj) ce qui témoigne de la difficulté à se plier à une description purement lexicale sans syntaxe. Les formes les plus fréquentes sont cependant conformes à celles du système.

L'analyse des SN apparus dans la formulation spontanée ou dans les mots clés produits par l'utilisateur ne s'avère donc pas une opération syntaxique simple dans l'optique d'un traitement automatique de la demande.

### 3.1.3. Les niveaux de généralité des descripteurs usagers et système

Pour comprendre les décalages de niveaux de spécificité du lexique entre usagers et système, il est nécessaire d'avoir en tête *le mode de gestion des classes dans le système* qui explique le choix de son niveau de base et que nous allons brièvement expliciter.

Les thésaurus (constitués des mots clés d'indexation) sont organisés selon une hiérarchie plus ou moins ample ; ces différents niveaux de généralité coexistent dans la zone mots clés de la BDB et l'on pourrait considérer que le niveau de base du système serait celui qui est choisi

pour définir les entrées de l'index. du produit imprimé (le premier niveau quand l'index est hiérarchisé par niveaux) ; c'est le niveau *choisi pour communiquer l'information*. On ne peut dire dans quelle mesure il exprime le modèle cognitif implicite que les indexeurs se font du lexique de l'utilisateur moyen. Le niveau de généralité de ces entrées d'index est, comme la richesse du thésaurus, le résultat de deux facteurs : la gestion extensionnelle des classes et l'étendue des domaines couverts par la BDB.

1) *La gestion extensionnelle quantitative des classes* : un mot-clé ne justifiera son existence que s'il subsume un nombre suffisant de références dans l'index annuel du produit imprimé. *C'est en effet la tâche de génération des index*, à laquelle est liée la norme du nombre de descripteurs par document, *qui commande la gestion en intension des classes, i.e. la décision d'introduire un nouveau descripteur*. Cette gestion quantitative annuelle des index entre en conflit avec le caractère cumulatif des bases de données. Les classes ainsi définies s'avèrent trop abondantes, non discriminantes sur plusieurs années et demanderaient à être spécifiées pour la BDB. On voit là en quoi le maintien en parallèle de deux tâches liées à deux outils de communication (indexation pour le produit imprimé, indexation pour la BDB) génère des conflits de gestion du niveau de spécificité du lexique et du sens.

2) Le niveau de généralité des classes dépend de *l'étendue des domaines couverts par la BDB* : plus ceux-ci seront étendus, plus les classes seront génériques (Tversky et Gati, in Rosch, 1978, p. 90<sup>5</sup>) :

---

<sup>5</sup> "...The diagnosticity of features is determined by the prevalence of the classifications that are based on them. Hence, features that are shared by all the objects under study are devoid of diagnostic value, because they cannot be used to classify these objects. However, when the context is extended by enlarging the object set, some features that had been shared by all objects in the original context may not be shared by all objects in the broader context. These features then acquire diagnostic value and increase the similarity of the objects that share them. Thus the similarity of pairs of objects in the original context is usually smaller than their similarity in the extended context."

Moins la BDB sera spécialisée, plus les traits à valeur diagnostique seront génériques. Or les BDB du système sollicité sont des BDB disciplinaires, dont le spectre est toujours plus large que celui de l'utilisateur : le système doit gérer une multiplicité de points de vue, ce qui n'est pas le cas de l'usager, qui omet l'existence des autres spécialités que la sienne. Le système est amené à gérer la multiplicité des concepts des différentes spécialités dans un lexique unique, et à simuler l'existence d'un langage et de concepts identiques entre spécialités. D'où la faible représentation de la conceptualisation théorique dans les thésaurus.

Ces deux facteurs, gestion extensionnelle des classes et étendue des domaines couverts laissent présager un écart important entre lexique des spécialistes (enseignants-chercheurs) et lexique du système. Qu'en est-il exactement ?

### *Décalages usagers/système concernant le lexique*

Comme pour l'analyse des SN, connaissant bien le système et son lexique nous avons procédé pour chacun des trois groupes (enseignants-chercheurs, professionnels et licence) à la traduction de 10 demandes de psychologie et sciences de l'éducation et de la demande en mots clés du système de la base sciences de l'éducation (voir exemples en Annexe 2). Il est alors possible de caractériser les transformations opérées dans cette traduction (cf. tableau 4 ci-dessous).

La première partie du tableau concerne les opérations d'appauvrissement du lexique de la demande, la dernière celles d'enrichissement.

*L'appauvrissement* concerne les requêtes d'enseignants-chercheurs, peu celles des autres groupes ; il consiste le plus souvent en suppressions de termes, parfois importantes, le service documentaire choisissant la stratégie de ne pas croiser le mot-clé selon les différentes facettes définies dans la requête (cf. plus haut : spécification et rappel) ; les suppressions d'adjectifs ou de noms sont parfois liées à des implicites du lexique du système, lié à un domaine : ex. : "résultats scolaires" sera traduit par "rendement".

Les *glissements conceptuels* consistent en l'établissement de synonymies non orthodoxes dans le langage scientifique ; ex. : "représentation de l'école" sera traduit par "attitude envers l'école" ; le remplacement d'un terme source par plusieurs termes cibles concerne essentiellement le lexique lié à des clés de tri contextuelles (les groupes d'âge et niveaux d'enseignement, davantage spécifiés par le système). Les licence sont les plus nombreux à émettre des *requêtes non traitables* car trop peu spécifiées (ex. : "la couleur

chez l'homme", "les théories de l'apprentissage"). On peut dire qu'il n'y a pas symétrie des groupes à forte et faible formation dans le domaine par rapport au système : alors que les Licence formulent souvent des questions trop génériques pour ce dernier, les questions formulées par les chercheurs sont acceptables même si la traduction, plus générique, entraîne du bruit ou une inorganisation des données en sortie. Par contre on peut noter que l'appauvrissement dans la traduction ne touche pas descripteurs et clés de tri contextuelles de la même façon : c'est la partie notionnelle, théorique (cf. en annexe l'ex. 157 sur les stéréotypes) qui est appauvrie, alors que les clés de tri contextuelles sont enrichies ; enfin, l'appauvrissement (et le bruit) sont liés à l'incapacité du système à prendre en compte la topicalisation des relations (ex. 157 : interaction entre motivation et cognition dans les tâches scolaires (aussi en fonction de la réussite et de l'échec), où le "ET" booléen n'est pas à même de traduire la notion d'interaction.

		Enseignants- Chercheurs	Profes- sionnels	Licence
Suppression d'1 T		9	5	5
Suppression de plusieurs T	2(25et7)	-	1	
Suppression Adj ds NmAdj	9	-	1	
Suppression de <u>Nm</u> ds <u>Nm</u> de Nm		2	1	2
Plusieurs TS pour 1 TC:	2	1	-	-
	3	1	-	-
	>4	2	1	-
T hors thésaurus		5	7	3
Remplacement par T +génér.		2	1	1
T° conceptuelle		3	3	1
Traitement impossible		-	1	7
T° nulle		14	19	18
Plusieurs TC pour 1 TS:	2	6	4	-
	3	2	2	1
	>4	1	1	-
Troncature		4	-	1
Ajout d'un SAUF		2	-	-
Ajout d'1 T syn.		2	-	1
Ajout de plusieurs T. syn.		2	-	-
Ajout d'1 T. non syn.		-	1	1
Remplacement par 1 syn.		3	6	4
Remplacement par T + spécif.		-	1	-

T= Terme ; TS= Terme source ; TC= Terme cible ; syn.= synonyme ; spécif.= spécifique ; T°= transformation

*Tableau 4 . Transformations opérées par la traduction de la requête dans le lexique du système*

On peut remarquer qu'en terme de structure des SN composant les descripteurs, le décalage entre professionnels ou Licence et système paraît moindre qu'entre E.C. et système. Les premiers, lorsqu'ils interrogeront avec leur lexique trouveront un équivalent dans le système, mais ces descripteurs appelleront un trop grand nombre de références (cf. plus haut, la gestion du lexique dans le système).

Du côté des chercheurs, le lexique du système paraît insuffisant pour les descripteurs (non pour les clés de tri contextuelles) ; la définition de requêtes spécifiées au niveau de base des chercheurs suppose l'usage de termes hors thésaurus.

On pourrait résumer la situation en disant que les niveaux de base des groupes en formation, des professionnels et du système sont plus proches entre eux que les niveaux de base chercheurs/système, mais que, à l'inverse de ce que se représentent les populations en formation (cf. début de ce chapitre), l'opération de généralisation (passage à un niveau plus générique) peut être plus facilement assurée par le service documentaire que l'opération inverse de spécification, ce qui provoquerait cette asymétrie chercheurs/groupes en formation par rapport au système.

### *3.2. Quelques inférences sur la représentation du système par les usagers*

#### *3.2.1. L'usage de "mots clés"*

L'objectif est d'analyser ce que les utilisateurs entendent par mot-clé, et s'ils en ont la même conception que le système. L'idée que le système procède à la recherche par mots clés peut être lié à une pratique de fichiers de bibliothèques ou d'index de bulletins bibliographiques mais n'implique pas une connaissance précise des normes du système.

*Usage des mots clés* : Sur les 150 requêtes analysées, 23 seulement font explicitement référence à des "mots clés", dont 10 enseignants-chercheurs, 5 DEA, 5 maîtrise, 2 professionnels, 1 licence.

On a soit de simples listes de termes séparés par une virgule ou un tiret, soit lorsqu'il n'existe pas de structuration explicite des listes, un ordre non aléatoire des mots clés : le

premier est le plus représentatif de la thématique (12 cas). Les mots clés disposés de façon linéaire sont en fait organisés en groupes sémantiques : l'exemple 43 (voir Annexe 2) montre le réseau associatif des mots clés.

On a aussi des listes hiérarchisées de façon explicite : par numérotation (un terme générique est suivi en retrait de ses termes spécifiques) ; hiérarchie domaine/mots clés ; hiérarchie nom de clé de tri contextuelle, liste des termes qui en relèvent. Deux exemples qui viennent de chercheurs montrent que mots clés et clés contextuelles sont bien différenciés dans leurs fonctions.

### *3.2.2. Opération de spécification, usage des opérateurs booléens et attentes concernant le rappel*

Certains sujets énoncent leur thématique de façon générique et se livrent ensuite à une définition très détaillée des différentes facettes de la question ; la représentation sous-jacente est que le rappel d'un terme sera d'autant plus riche qu'il sera croisé avec un plus grand nombre de termes (ou l'extension de A sera inférieure à l'extension de  $(A * (B1 \text{ OU } B2 \text{ OU } \dots Bn))$ ) ; cette représentation erronée est partagée par les personnes qui n'ont pas l'expérience des systèmes booléens. Mais il n'est pas exclu que la définition d'une stratégie complexe de croisements vise l'organisation sémantique des références en sortie : il peut s'agir d'organiser ces dernières selon des points de vue ou approches, ce que le système ne prévoit pas et qui pourrait aider à la lecture et à la sélection dans les listings.

Deux sujets seulement utilisent explicitement des opérateurs booléens et disposent graphiquement les groupes en vue de la combinatoire. Un sujet utilise le terme "croiser" équivalent du "ET" booléen. On voit donc que la définition opératoire de la requête sans intermédiaire humain reste très diverse selon les usagers et son décodage automatique par le système pose de multiples problèmes.

### *3.2.3. Thème de recherche, thème de la bibliographie et mouvements de paraphrasage*

On fera l'hypothèse que la différenciation du thème de travail et du thème de la bibliographie témoigne d'un effort chez l'utilisateur pour *reformuler la demande en fonction de son modèle mental du système*. Comme le confirme le tableau ci-dessous, le thème de recherche présenté seul sans spécifications pour la bibliographie diminue avec le niveau de formation ; les groupes en formation formulent plus souvent que les enseignants-chercheurs et les professionnels leur demande en deux (ou trois) temps : thème de recherche/thème de la bibliographie et le thème de recherche est plus souvent placé en premier.

## Thème de la bibliographie

		+					-					
		Licen	Maît	DEA	E.C.	Prof	Licen	Maît	DEA	E.C.	Prof	
Thèm Rech.	+	<b>1</b>	77,7	70,3	60,6	30,3	21,2	55,5	26	24,2	12	10,5
		<b>2</b>	5,5	3,7	21,2	-	-					
		<b>3</b>	-	-	12,1	-	-					
	-		16,6	26	15,1	66,6	76,3					

Tableau 5. Fréquences d'occurrence en % sur les colonnes du thème de recherche en position 1, 2, ou 3 par rapport au thème de la bibliographie

Précisons que les marqueurs du thème de recherche sont les expressions du type : "je travaille sur", "mon thème de recherche porte sur", "ma thèse porte sur", "je m'intéresse à" ; les marqueurs du thème de la bibliographie sont du type : "je cherche des documents sur", "je souhaite une bibliographie sur".

*a) Les mouvements de paraphrasage du thème de recherche vers le thème de la bibliographie*

Dans le corpus issu des demandes par correspondance, il est difficile de démêler ce qui, dans la formulation de la demande de bibliographie relève des processus d'organisation textuelle et ce qui relève du modèle mental du système.

On peut en effet supposer que, comme dans toute organisation textuelle, se formule d'abord l'expression générique du thème (souvent représentée par un titre) et dans la suite du paragraphe sont formulées les spécifications du thème générique (Kintsch, 1990) ; ces spécifications peuvent être données à l'intérieur du thème de recherche ou dans le thème de la bibliographie. Ce dernier apparaissant souvent en deuxième position, serait porteur de ces spécifications ; si, par contre, on observe un mouvement de généralisation (passage à du plus générique), qui est contraire à l'organisation textuelle hiérarchique, ce mouvement devrait être lié au modèle mental du système.

*b) Les mouvements du thème de recherche en position 1 vers le thème de la bibliographie en position 2 (70 cas)*

On observe deux fois plus de spécifications que de généralisations (36 contre 17) et 6 reformulations identiques. Les opérations de spécifications concernent surtout des spécifications de concepts et ajouts de mots clés (16 cas), des ajouts de clés contextuelles (population, période, genre littéraire, etc. : 8 cas), l'ajout du point de vue disciplinaire (4

cas) ou de partie d'un script (3 cas), enfin, déplacement d'une clé contextuelle en descripteur (2 cas).

Les opérations de généralisation portent surtout sur la suppression de clés contextuelles (7 cas) et l'usage de mots clés plus génériques (6 cas), la suppression de descripteurs (3 cas) qui sont des concepts abstraits portant sur des processus : ainsi, la notion de "représentation", source d'embarras dans la formulation de la requête est délibérément supprimée par l'utilisateur dans sa reformulation (ce que nous interprétons comme une difficulté à lexicaliser une notion abstraite). De même, le point de vue est supprimé deux fois.

Il est à noter que les généralisations n'existent pratiquement pas chez les enseignants-chercheurs, et se produisent chez les maîtrise et DEA, alors que les spécifications se produisent dans tous les groupes sauf chez les professionnels, ce qui laisse supposer des images différentes du système.

Lorsque *le thème de recherche apparaît en deuxième ou troisième position* (11 cas seulement qui ne concernent que les populations en formation) les transformations observées consistent en spécifications lexicales (spécifications ou ajouts de mots clés (6 cas), ajout de clés contextuelles (4 cas), ajout de clés non thématiques (2 cas), mais aussi formulation sous forme de phrases verbales de la problématique, d'hypothèses, d'analogies et spécification de relations entre concepts (6 cas).

Les opérations de généralisation portent sur les clés contextuelles (4 cas) et la formulation d'un titre générique (2 cas), ce qui laisserait supposer que le titre générique (forme discursive définissant le thème de travail) joue le rôle de "bonne forme" par rapport à une "mauvaise forme", purement lexicale donnée par une liste de termes.

Sur l'ensemble des transformations, quel que soit le thème apparu en premier, les types les plus fréquents (74% des transformations) portent sur les descripteurs (ajout et spécification de termes : 39,4%) et sur les clés contextuelles (33,8%). On voit donc que les clés de tri contextuelles occupent une place importante dans les transformations lexicales, et ceci davantage dans les généralisations (45,8% de ces dernières) que dans les spécifications (27,8% de celles-ci) : la reformulation s'avère plus aisée sur les parties moins conceptuelles de la requête.

De plus le groupe qui utilise le lexique le plus spécifique ne généralise pas, et la tendance à généraliser chez les maîtrise et DEA laisse supposer un modèle trop générique du lexique du système.

#### ***IV. Conclusion***

En amont des problèmes de gestion du conceptuel à travers le linguistique, et dans une optique de compréhension automatique de la demande, on a pu noter combien l'analyse linguistique soulève de



problèmes divers sans compter les fonctions métadiscursives de la disposition graphique des textes de demande.

Nous pensons avoir mis en évidence l'hétérogénéité de la demande. La variabilité des distances des usagers au système concernant à la fois l'usage des différents critères de sélection et les niveaux de généralité du lexique peut être résumée ainsi :

1) L'analyse des critères de sélection a montré : i) l'usage de critères absents du système (touchant aux genres textuels et à la différenciation de facettes et clés de tri contextuelles ; ii) la sous-utilisation générale de clés de tri non-thématiques, plus marquée pour les populations les plus distantes de l'institution universitaire et de ses modes de catégorisation des productions écrites. Les clés de tri thématiques, elles, s'enrichissent avec le niveau de formation.

Distinguer les trois éléments constitutifs du schéma de base de la demande de bibliographie en sciences sociales (point de vue disciplinaire, objet d'étude et clés contextuelles) nous a permis de dégager les fonctions cognitives différenciées de ces trois éléments dans le fonctionnement de la demande et des différences de spécification selon les groupes : la spécification du point de vue sur l'objet s'avère particulièrement importante pour la navigation dans les BDB (cf. les travaux de définition de pré-serveurs), et non trivial car l'interrogation suppose non seulement de repérer les disciplines pertinentes, mais aussi d'éliminer les spécialités non-pertinentes, ce qui demande un décentrage par rapport à son point de vue propre. Ceci est d'autant moins trivial que la désignation des spécialités par le système et par les communautés scientifiques présentent un fort décalage. Quoiqu'il en soit, le repérage disciplinaire s'avère plus problématique pour les populations les moins imprégnées des découpages académiques. Il est clair que le degré de spécification du point de vue aura des conséquences précises sur l'interprétation de la demande par le système et les choix lexicaux liés aux descripteurs de l'objet d'étude (cf. infra).

Quant au statut différent des clés contextuelles et des descripteurs, on peut considérer qu'à l'intérieur du schéma de base, les clés contextuelles correspondent à la partie stable, consensuelle, à un savoir encyclopédique de base ("context-independent information" selon Barsalou, 1989) partagé par l'ensemble des usagers et par le système, alors que le signifié des descripteurs, lié au point de vue sur l'objet et au niveau d'expertise, ne sera pas forcément homogène entre sujets et système. L'usage multi-spécialités d'une même terme par le système et le caractère contrastif des termes choisis comme entrée dans le thésaurus génère dans l'usage du lexique des ambiguïtés et des normes d'emploi internes au système qui ne pourront être repérées qu'en contexte par l'utilisateur.

2) Nous avons montré que le système rencontre des difficultés dans le traitement des questions émanant des moins formés et des plus formés (enseignants-chercheurs), avec une impossibilité fréquente à répondre aux questions trop génériques des premiers. Nous avons vu par l'étude de la structure syntaxique des descripteurs que ces derniers sont plus spécifiques chez les sujets plus formés et donc sans doute moins ambigus. Une *modélisation cognitive* de l'utilisateur à travers les traces verbales de la demande devrait pouvoir s'attaquer au diagnostic du type de questionnement épistémique sous-jacent à la demande (Dillon 84), qui devrait aboutir à des routines de guidage différenciées selon les niveaux de formation, portant en particulier sur les genres d'élaborations (ouvrage de base, trend reports) et sur les domaines de recherche appelés par la requête. Concernant le niveau de généralité choisi pour le lexique, cette étude exploratoire suggère que la gestion de ce dernier dans le système est en décalage avec le lexique scientifique des spécialistes : son manque de spécificité lié à l'étendue des domaines couverts par la BDB et la gestion des classes en extension sont à l'origine de difficultés à l'interrogation.

Le problème du choix du mode de représentation du texte a été souligné par l'analyse de la structure syntaxique des mots clés : si la conception du mot-clé chez les utilisateurs reconnaît bien la prévalence du SN qui est la règle dans le système, elle a du mal à se mouler dans sa conception stricte et accepte des structures syntaxiques beaucoup plus complexes que ce que le système admet : chassez la discursivité, elle revient au galop. Ceci met en question la validité d'une description purement lexicale des textes et de la demande et invite à passer à une description en *frames* plus riche qu'une liste d'attributs et comportant les relations entre catégories.

Une étude linguistique des processus de spécification et généralisation reste à faire, en parallèle avec l'étude cognitive des rapports concrétude/abstraction (Lakoff, Medin 1989) dans l'interprétation textuelle : si, chez l'utilisateur le schéma a priori de l'arborescence monohiérarchique domine, lui faisant adopter une stratégie économique de formulation générique qui appellerait tous les spécifiques (schéma qui vaut peut être pour les clés contextuelles), il paraît clair que l'opération de généralisation comme de spécification ne fonctionne pas pour les concepts abstraits, en monohiérarchie et dépend de la gestion des points de vue sur l'objet (d'où l'importance d'un repérage fin des domaines de recherche et des cadres théoriques). Il est patent que dans sa conception référentielle du langage ("un objet, un nom" est translaté à un "concept, un nom") l'organisation classificatoire des BDB (plan de classement et thésaurus) privilégie les objets concrets, compatibles avec cette organisation quasi-taxinomique, alors que la gestion des concepts théoriques, qui ne répondent pas à cette organisation hiérarchique stable s'avère beaucoup plus problématique pour le système. Ceci se vérifie dans l'importance attribuée par ce dernier aux clés contextuelles pour gérer la spécification ou l'élargissement des questions à l'interrogation. Il faut de plus noter que le jugement de pertinence d'un document dépend de l'accessibilité à des données concrètes opératoires (en particulier

description des procédures) qui permettent les inférences les plus riches sur l'opérationnalisation des concepts, données souvent négligées par les BDB. Tout cela conduit à des choix concernant la représentation du texte dans le système différents de ceux du système analysé ici. Le développement de l'hypertexte et de l'accès au texte intégral orientera le travail vers la mise au point d'outils de reconnaissance des types de textes et, à l'intérieur, des fonctions des différentes parties d'un texte scientifique et à la prise en compte de la dimension argumentative et dialogique du discours scientifique.

Denise MALRIEU  
CNRS-CAMS  
96 Bd Raspail, 75006 Paris

## Annexe 1

### *Les contraintes du système sollicité : traitement et logiciel d'interrogation*

Nous donnerons ici un bref descriptif des caractéristiques du système sollicité :

- Les types de documents traités : ouvrages et compte rendus d'ouvrages, articles de revues scientifiques, rapports de recherche, thèses sont les 5 types de documents retenus par le système.
- Les zones interrogeables : nom de la revue, date de parution, localisation du document, noms d'auteurs, titre, résumé, mots clés d'indexation, plan de classement (grands chapitres de la base de données). Selon les serveurs et leurs logiciels d'interrogation, la recherche de termes hors thésaurus dans le titre et le résumé peut ou non se faire avec des indicateurs de proximité : termes adjacents, distants de  $n$  mots, dans l'ordre indiqué ou dans un autre ordre, appartenance au même champ.
- Les spécifications obligatoires pour l'interrogation (elles apparaissent sur le bordereau de commande de bibliographie automatisée) : nombre d'années à interroger ; langues des documents ; nombre maximum de références souhaité, ou coût maximum de l'interrogation, nom de la ou des bases à interroger (chacune correspondant à une discipline) ; aires géographiques concernées.

---

## Annexe 2

Voici deux exemples de lettres:

1 - *D'enseignant-chercheur* (n° 157):

Chère Madame,

Je me permets de vous écrire pour préciser notre demande bibliographique. Comme je vous l'ai signalé, il s'agit de 4 thèmes:

### 1) Stéréotypes (sociaux et ethniques)

A propos de ce sujet on souhaiterait en priorité obtenir les revues de question à partir de 1980.

Nous cherchons des études expérimentales qui illustrent le découpage selon différents niveaux d'explication:

- niveau intraindividuel (formation des impressions, sélection de l'information, structuration et traitement de l'information, effet d'assimilation et de contraste).

- niveau interindividuel (hiérarchie organisationnelle, hiérarchie institutionnalisée, réseaux hiérarchisés, rapports de pouvoir à l'intérieur d'une organisation, d'une institution ou d'un groupe).

- niveau positionnel (différentiation selon le sexe, selon l'appartenance socioéconomique, selon l'appartenance ethnique).

- niveau idéologique : (différenciation selon l'appartenance politique, syndicale, religieuse ; orthodoxie). Nous sommes également intéressés par tous les travaux concernant les fonctions aussi bien cognitives que sociales des stéréotypes.

Mots clés du système : stéréotype ; (relations ethniques, relations interpersonnelles, relations raciales, relations intergroupes, relations humaines, relations de travail) ET (perception d'autrui, perception de rôle, attente de rôle)

(.....)

4) interaction entre la motivation et la cognition dans les tâches scolaires (aussi en fonction de la réussite et de l'échec) ; études longitudinales.

Mots clés du système : Motivation/, /aspiration, renforcement/ ; /cognition/ ; apprentissage ; compréhension ; mémorisation ; sauf apprentissage social ; résolution de problème ; élève du primaire, élève du secondaire.

### 2 - D'étudiant :

Monsieur,

Préparant dans le cadre de l'Université de ... un troisième cycle dont le projet de recherche porte sur l'image de soi chez l'adolescent socialement inadapté, je vous serais reconnaissant de bien vouloir me fournir une bibliographie exhaustive et actuelle des ouvrages et articles, français et étrangers, parus à ce jour et traitant de ce thème.

*Exemple de liste de Mots clés (n° 43) :*

Anorexie - boulimie - oralité - culture - tradition arabo-musulmane - pratiques alimentaires - éducation dans les traditions arabo-musulmanes - famille et société algérienne - psychanalyse et culture - ethnologie - ethnopsychologie - ethnopsychiatrie - ethnopsychanalyse - test projectif dans les cultures - maigreur et minceur dans les différentes cultures - obésité - introjection - projection - identification - assimilation

culturelle - famille et société - famille et culture - conflits de cultures - conflits de générations - culturation - acculturation - déchirement culturel - enracinement - le point de vue culturel des maladies mentales et troubles psychologie - psychopathologie - éléments psychosociaux dans les troubles alimentaires - rituels alimentaires.

## *Bibliographie*

- BARSALOU L.W. (1989) Intraconcept Similarity and its Implications for Interconcepts Similarity, in Vosniadou & Ortony : *Similarity and Analogical Reasoning*. Cambridge University Press, 76-121.
- BELKIN N.J., ODDY R.N., BROOKS H.M. (1982) Ask for information retrieval, *J. of Documentation*, 38, 2 et 3, 61-71 et 145-164.
- BOURION E. , MALRIEU D. (à paraître) Etude sémantique et sémiotique d'un plan de classement de banque de données : concepts et systèmes signifiants.
- BROOKS H. M., DANIELS P.J., BELKIN N.J. (1985) Using problem structure for driving human-computer dialogues. *RIAO*, 645-660.
- CARBERRY S. (1988) Modeling the User's Plans and Goals. *Computational Linguistics*, 14, 3, 23-37.
- CHEN H., DHAR V. (1990) User misconceptions of information retrieval systems. *Int. J. Man-Machine Studies*, 32, 673-692.
- CHIARAMELLA Y., DEFUDE, B. (1987) IOTA : Un prototype de système expert en recherche d'informations, *7èmes Journées systèmes experts d'Avignon*, 511-528.
- COQUET J.-C. (1984) *Le discours et son sujet*, Paris, Klincksieck.
- CRUSE D.A. (1977) The pragmatics of lexical specificity, *J. of Linguistics*, 13, 153-164.
- DILLON J.T. (1984) The classification of research questions, *Review of Educational Research*, 54, 3, 327-361.
- DUBOIS D. (1991) Catégorisation et Cognition : "10 ans après" une évaluation des concepts de Rosch in D. Dubois, éd., *Sémantique et Cognition. Catégories, prototypes, Typicalité*. Paris, Editions du CNRS.
- FALZON P. (1989) *Ergonomie cognitive du dialogue*. Grenoble, Presses Universitaires de Grenoble, Collection Sciences et Technologie de la connaissance.
- KINTSCH E. (1990), Macroprocesses and microprocesses in the development of summarization skills. *Cognition and Instruction*, 7, 3, 161-195.
- MALRIEU D. (à paraître) Taxinomies ou schémas ? L'organisation des catégories dans les plans de classement des banques de données bibliographiques. GDR Sciences Cognitives de Paris, Atelier Catégories, concepts et systèmes symboliques, Journées de Brest 1991.
- MALRIEU D. (à paraître) Approche pragmatique du discours de demande bibliographique et modélisation de l'utilisateur.
- MARTIN T.H., CARLISLE J., TREU S. (1973) The user interface for interactive bibliographic searching : an analysis of the attitudes of nineteen information scientists. *J. of the American Society for Information Science*, 142-147.
- MCCOY K.F. (1989) Generating Context sensitive responses to object related misconceptions, *Artificial Intelligence*, 41, 157-195.



- MEDIN D. (1989) Concepts and Conceptual Structures, *American Psychologist*, 44, 12, 1469-81.
- PARIS C. L. (1988) Tailoring object descriptions to a user's level of expertise. *Computational Linguistics*, 14, 3, 64-78.
- PEJTERSEN A.M., AUSTIN J. (1983) Fiction retrieval : experimental design and evaluation of a search system based on user's values criteria, *J. of Documentation*, 39, 4, 230-246.
- POLLITT A. S. (1987) CANSEARCH : An expert systems approach to document retrieval, *Information Processing & Management*, 23, 2, 119-138.
- PROTÉE (1988) 16, 1-2, n° spécial : Le point de vue fait signe.
- QUILICI A., DYER M., FLOWERS M. (1988) Recognizing and responding to plan-oriented misconceptions. *Computational Linguistics*, 14, 3, 38-51.
- RASTIER F. (1987) *Sémantique interprétative*, Paris, PUF.
- RASTIER F. (1988) Sémantique de l'interaction verbale Homme/Machine, in Cosnier J. et al. (Ed.) *Echanges sur la conversation*, Lyon, CNRS, 301-322.
- RASTIER F. (1989) *Sens et textualité*, Paris, Hachette.
- ROSCH E. (1976) Basic objects in natural categories, *Cognitive Psychology*, 8, 382-439.
- ROSCH E. (1978) Principles of categorization, in E. Rosch & B.B. Lloyd (Eds), *Cognition and Categorization*, N.J. : Lawrence Erlbaum Assoc.
- ROUAULT J. (1987) *Linguistique informatique, Applications documentaires*, Berne, P. Lang.
- SHOVAL P. (1986) Comparison of decision support strategies in expert consultation system. *International J. of Man-Machine Studies*, 4, 2, 125-139.
- SOWA J. (1984) *Conceptual structures*, Reading, (Mass.) Addison-Wesley.
- TANAKA J.W., TAYLOR M. (1991) Object categories and expertise : Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 3, 457-482.