Céline Paganelli
Evelyne Mounier
GRESEC, Université Stendhal Grenoble III
Institut de la Communication et des Médias
11, Avenue du 8 mai 1945
38130 Échirolles

Influence des variables métier et expertise dans l'accès à l'information dans les documents professionnels : application à la documentation d'une entreprise du secteur bancaire

Résumé : Cette communication porte sur la recherche et l'accès à l'information professionnelle contenue dans les documentations d'entreprise. Une étude expérimentale menée au sein d'une banque montre que la disparité entre le vocabulaire utilisé dans les documents professionnels et celui employé par les différents personnels de l'entreprise induit des difficultés dans l'accès à l'information.

Abstract: This communication concerns information retrieval in the professional documentations of a banking company. An experimental study shows differences between the vocabulary contained in these documents and the vocabulary used by the employees of this company. These disparities imply difficulties in the access to information.

1. Introduction

La place de l'information dans l'entreprise est aujourd'hui prépondérante. Les personnels en situation de travail sont à la fois producteurs d'information en rédigeant par exemple des documents utiles au fonctionnement de l'entreprise, utilisateurs d'informations et de documentations, en passant notamment par les circuits d'information crées au sein de la structure, et enfin gestionnaires d'information, par le biais de l'organisation et du classement des informations et documents qui leur sont utiles, mais aussi en développant par exemple des systèmes d'information parallèles et qui leur sont propres.

Cette communication présente les résultats d'une étude effectuée au sein d'une banque. Elle analyse le comportement informationnel du personnel d'une entreprise du secteur bancaire, et porte plus précisément sur la recherche de documents et l'accès à l'information professionnelle contenue dans les documentations internes à l'entreprise.

2. Contexte de l'étude

La gestion rationnelle des informations et des documents dans les entreprises amène souvent celles-ci à porter sur format électronique l'ensemble des documents qu'elles produisent de manière à en améliorer l'accès pour leurs employés.
Sur support électronique, ces documentations professionnelles constituent des bases de données dans lesquelles on constate :

- que la version papier a été, le plus souvent, transposée telle quelle sur le support électronique,

- que les documents regroupés sont très hétérogènes : on retrouve aussi bien des documents peu structurés (emails, notes de service, etc.) que des documents très normalisés (documentations techniques par exemples),

L'accès aux documents ou aux informations pertinentes est lié à l'efficacité des recherches dans ces bases de données ; or, elle peut être affectée par de très nombreux facteurs qui tiennent aussi bien aux performances du système d'information qu'aux caractéristiques des usagers. Leur niveau de connaissance par rapport au thème de la recherche d'information est un de ces facteurs. Par exemple, un expert d'un domaine produit globalement peu de requêtes, mais ces dernières sont souvent plus pertinentes que celles produites par un novice dans le même domaine [MARCHIONINI, 1991] ; de même, les experts d'un domaine comprennent assez vite le problème à résoudre, sont capables d'anticiper et de prédire les réponses proposées par le système, n'utilisent que des termes de recherches très spécifiques pour produire leurs requêtes, consacrent beaucoup de temps au traitement des résultats, et surtout sont capables d'évaluer la pertinence de ces résultats de manière relativement rapide.

Dans ces systèmes d'informations, les modes de consultation proposés correspondent à ceux que l'on trouve habituellement dans la consultation des bases de données ou dans les moteurs de recherche en texte intégral. Pourtant, la recherche d'information dans les documentations professionnelles présente des particularités :

- on peut y chercher la réponse précise à une question, par exemple en vue de dépanner une machine ou un logiciel, mais on peut aussi « feuilleter » parcourir une documentation, ou encore la lire in-extenso « pour voir comment cela fonctionne ».
- les utilisateurs effectifs de ces documents sont toujours en situation de travail, ce qui induit d'une part que l'information cherchée est nécessaire à la prise d'une décision ou à la résolution d'une difficulté technique, et d'autre part, l'existence d'une contrainte de temps.

On peut donc considérer que la recherche d'information dans les documents professionnels est guidée par un ensembles de facteurs tels que : l'objectif de l'utilisateur, ses connaissances sur le domaine, la représentation qu'il se fait du système d'information, ses fonctions, le travail qu'il doit exécuter, etc. Si le but de l'utilisateur est d'accéder, aussi directement que possible, à l'information précise et utile, il doit rechercher mais aussi évaluer l'information dont il a besoin.

Un autre aspect important de la recherche d'information dans les documents professionnels est donc celui de l'évaluation de la pertinence des résultats. Depuis les premiers travaux dans les années 80, un très grand nombre de critères ont été identifiés. Par exemple, selon Bruce [BRUCE 1994], les adultes utilisent deux types de critères : des critères objectifs (tels que les titres, mots-clés, noms des auteurs, date de publication) et des critères dits subjectifs comme la précision des informations, leur intérêt, le coût cognitif associé au traitement, l'exactitude des informations, etc. Cependant, il semblerait que tous ces critères interviennent de manière différente. Ainsi, Peiling & Sorgel [PEILING, 1998] ont montré que :

- *les titres* (des ouvrages ou des pages Web) ainsi que les résumés servent surtout à évaluer la *pertinence thématique* du document.
- les noms d'auteurs et les noms de la collection sont des indices qui servent surtout a évaluer la *qualité du document et* la nouveauté du document

Néanmoins, nous remarquons que l'ensemble des travaux concerne plutôt la recherche de documents que la recherche d'informations, par ailleurs, on connaît moins bien les

facteurs qui affectent les jugements de pertinence dans le monde du travail. Ainsi, une autre manière d'envisager la pertinence de l'information est de la considérer comme liée à son utilisabilité [HARTER 1992 ; LECOADIC 1997]. Dans ce cas, une information est considérée comme pertinente à condition qu'elle respecte deux conditions : elle doit être utile pour l'utilisateur et elle doit être utilisable facilement.

A l'heure actuelle, l'accès aux documents d'entreprise et à l'information professionnelle se heurte donc à un problème complexe :

- D'une part, comme on l'a déjà dit, ces documents servent de bases d'informations dans lesquelles les professionnels effectuent leurs recherches d'informations en utilisant de plus en plus souvent des requêtes en langue naturelle. Cependant, les utilisateurs ne retrouvent pas toujours les documents utiles,
- D'autre part, différents corps de métiers au sein d'une même entreprise, peuvent être amenés à consulter et à utiliser une même documentation professionnelle, notamment lorsque celle-ci contient des règles internes ou des dispositions administratives qui concernent l'ensemble de l'entreprise. Or, on constate une situation paradoxale : certains documents jugés indispensables par les informations véhiculées, sont également des documents peu ou pas utilisés par les utilisateurs potentiels.

De fait, pour qu'un document professionnel soit retrouvé, et qu'il soit exploitable par les utilisateurs auxquels il est destiné, il faut :

- D'une part que le vocabulaire utilisé dans les documents professionnels soit proche de celui qui est employé par les utilisateurs auxquels ces documents sont destinés ;
- Et d'autre part que le langage d'indexation le permette et notamment que le langage d'interrogation proposé, que ce soit un langage documentaire ou le langage naturel soit compris des utilisateurs visés.

Nous montrons ici que les documentations professionnelles sont peu utilisées parce qu'elles véhiculent des connaissances, des représentations mentales de ceux qui les ont rédigées et parce que le vocabulaire employé pour décrire une même tâche, varie d'un métier à l'autre mais aussi selon le niveau d'expertise dans un même métier. Elles sont également difficilement interrogeables via les requêtes en langue naturelle pour les mêmes raisons.

Parmi les sources d'information disponibles dans l'entreprise bancaire qui fait l'objet de cette étude, on trouve des documents réglementaires ou normatifs, élaborés au sein même de l'entreprise et jugés comme documents incontournables et de référence par leurs concepteurs. Ces documents recensent, de manière exhaustive, les règles à mettre en œuvre et à respecter dans le cadre du fonctionnement courant de la banque. Les collaborateurs de la banque, quelle que soit leur fonction, estiment eux-aussi que ces documents sont des sources de référence, riches d'informations et indispensables dans la gestion de leur activité. Il apparaît pourtant que ces documents sont très inégalement utilisés : la fréquence de consultation des documents dépend des fonctions assumées par les collaborateurs de l'entreprise. Ainsi, les contrôleurs, qui ont à faire respecter les règles administratives, sont ceux qui les consultent le plus souvent.

Mais d'une façon générale, tous les sujets rencontrés s'accordent pour dire que leur consultation n'est pas aisée. Plusieurs raisons sont évoquées : il existe d'abord un problème de repérage des documents pertinents, mais aussi un problème d'accès à l'information à l'intérieur du document et enfin de compréhensibilité du document.

Le système d'information permet plusieurs types de recherche : via un thesaurus spécialisé, via un plan de classement des documents, via des requêtes en langue naturelle. D'une manière générale, le thésaurus existant n'est pas utilisé pour l'interrogation. Les deux derniers modes d'interrogation sont les plus pratiqués par les usagers : quelques uns, familiers avec le mode de classement préfèrent ce mode d'accès ; les autres, que le plan de classement rebute, formulent des requêtes en langue naturelle mais ils ont du mal à atteindre les documents pertinents.

De plus, atteindre un document utile et pertinent n'est pas suffisant, encore faut-il qu'il soit exploitable par l'usager, ce qui pose un problème de lecture et de compréhension : ainsi ces documents ne sont pas conçus pour une lecture sur écran, ils sont beaucoup trop longs, les titres dans les tables des matières ne sont pas toujours clairs et explicites, la navigation à l'intérieur du document est difficile (l'accès aux annexes directement à partir du texte n'est pas possible). En dehors de l'aspect de conception des documents qui a été étudié par ailleurs, il reste celui de la compréhension du vocabulaire utilisé dans les documents. D'ailleurs, les utilisateurs considèrent qu'il faut des connaissances et de l'expérience pour arriver à comprendre, interpréter et exploiter l'information contenue dans les documents. De ce fait, ce sont les utilisateurs les plus expérimentés qui les utilisent le plus régulièrement.

3. Hypothèses

Nous considérons que le système d'information ne permet pas aux usagers de retrouver les documents et les informations utiles parce que la documentation visée correspond seulement aux connaissances, aux besoins et aux représentations d'une catégorie d'utilisateurs, celle qui rédige la documentation. Ainsi, ces différences doivent être perceptibles via :

- le vocabulaire employé qui devrait différer selon l'expertise des individus, notamment dans la description d'une même tâche ou d'un même problème.
- les besoins exprimés, qui doivent varier en fonction des tâches qu'ils remplissent habituellement. Certains types de documents peuvent se révéler peu utiles ou peu adaptés selon les tâches remplies.
- Le classement systématique des documents qui ne correspond pas non plus aux représentations des employés.

4. Méthodologie

Une tâche pilote a été choisie ; cette tâche est désignée par le terme « d'Entrée en Relation ou EER. Elle est à la base de toute l'activité des agents d'une entreprise bancaire. Elle recouvre à la fois l'accueil des clients réguliers, la prise de contact avec les nouveaux clients, l'ouverture d'un dossier d'ouverture de compte, les demandes éventuelles de prêts, la vente de services et produits bancaires. Elle est aujourd'hui très encadrée et contrôlée par le système d'information. Par ailleurs, la législation française a rendu les entreprises bancaires responsables devant la loi, des opérations et des transactions qu'elles acceptent pour leurs clients. De ce fait, ces entreprises sont amenées à durcir les mesures de contrôles. La formation et l'information des employés de l'entreprise sont donc primordiales, de même que les documents servant de référence qui doivent être mis à jour très régulièrement et être accessibles à tous.

On étudie d'une part, comment les différents corps de métiers décrivent cette même tâche, et d'autre part, les requêtes que les employés sont amenés à formuler quand ils se

posent des questions au sujet de cette activité ; on étudie également comment ils recherchent et obtiennent ou non l'information souhaitée. Deux expériences ont été menées auprès de sujets employés dans l'entreprise, assurant des fonctions différentes. Pour chaque fonction, 2 niveaux d'expertise ont été retenus : novices ou experts.

L'analyse porte d'abord sur le vocabulaire recueilli auprès des différents corps de métiers et ensuite sur un corpus de questions réelles entraînant des recherches d'informations mais aussi des regroupements d'informations.

4.1. Expérience 1 : Description de la tâche d'Entrée en Relation (EER)

Objectif

L'objectif de cette première expérimentation est de faire expliciter par une personne une série d'actions plus ou moins planifiées, de manière à recueillir le vocabulaire utilisé pour décrire une tâche professionnelle courante mais aussi pour essayer de comprendre quels sont les besoins récurrents en information, comment ces besoins en informations sont formulés, et quelles sont les pratiques habituelles mises en œuvre pour obtenir l'information utile. La tache décrite ici est l'entrée en relation avec le client (EER).

Méthode et sujets

Une première méthode a été utilisée, basée sur la technique d'entretien qui consiste à utiliser les questions « comment ? » et « pourquoi ? », pour inciter l'utilisateur à décrire une tâche donnée.

Mais il s'est révélé rapidement impossible d'appliquer *stricto sensu* cette méthode. Cette difficulté semble être liée à la simplicité (relative) de la tâche d'EER ou au fait que cette tâche est peu structurée hiérarchiquement. Une grille d'entretien a donc été élaborée, de manière à structurer les échanges. Outre les aspects liés à la description de la tâche, elle comprend des questions sur les problèmes d'information qui peuvent se poser aux interviewés, que ce soit des problèmes directement liés à l'EER ou non, et sur la manière dont les interviewés retrouvent l'information permettant de répondre à ces problèmes.

Pour cette première expérimentation, 18 sujets ont été interviewés. Les entretiens ont duré en moyenne une heure. Ces sujets se répartissent 3 groupes professionnels :

- Les commerciaux sont affectés à l'accueil de la clientèle. Ce sont eux qui, la plupart, du temps répondent aux questions des usagers et procèdent à l'ouverture des comptes.
- Les animateurs commerciaux peuvent avoir des responsabilités de service ou d'agence.
- Les contrôleurs exercent leur activité en « back office ». Ils sont chargés de vérifier la validité des dossiers, d'intervenir dans les cas litigieux.

Le tableau ci-dessous récapitule le nombre de sujets en fonction de leur métier et de leur niveau d'expertise.

	Commerciaux	**Animateurs Commerciaux**	**Contrôleurs**	**Total**
Experts	3	4	5	12
Novices	4	1	1	6
Total	7	5	6	**18**

Tableau 1 : Nombre de sujets par catégories et par niveau d'expertise

Analyse

Les principaux résultats obtenus sont les suivants : La représentation que les utilisateurs se font de la tâche et vocabulaire sur la tâche, la comparaison entre les différentes catégories d'utilisateurs, et la réalisation d'une liste de demandes d'informations potentielles, qui doit servir de support pour l'expérience suivante.

4.2. Expérience 2 : Tri et catégorisation des unités d'information

Objectif

L'objectif est de mettre à jour rapidement comment les experts d'un domaine regroupent un ensemble d'objets de leur domaine en différentes catégories.

Procédure et sujets

Nous avons utilisé la technique de tri [BISSERET 1999] qui consiste à présenter les unités d'information (issues de l'expérimentation précédente) en désordre et à demander au sujet de les trier en plusieurs groupes distincts tels que les objets du même groupe aillent bien ensemble. Lorsqu'un tri est fait, on interroge le sujet sur ce qui différencie, pour lui, les groupes. On lui demande également d'attribuer une dénomination pour chaque groupe.

9 sujets ont été interviewés. Ils se répartissent de la manière suivante :

	Commerciaux	**Animateurs Commerciaux**	**Contrôleurs**	**Total**
Experts	3	2	2	7
Novices	1	1	0	2
Total	4	3	2	**9**

Tableau 2 : Nombre de sujets par catégories et par niveau d'expertise.

Analyse

Cette expérimentation nous a permis de recueillir les catégories réalisées par les sujets, et d'évaluer les similitudes et des différences entre les regroupements.

5. Résultats

5.1. Résultats de la première expérimentation

Les entretiens réalisés au cours de la première expérimentation ont permis d'obtenir des informations autour des points suivants : la représentation que les sujets se font de la tâche d'entrée en relation, le vocabulaire qu'ils utilisent pour décrire cette tâche, et les problèmes d'information qu'ils peuvent rencontrer lors de la réalisation de cette tâche et qui donnent lieu à des demandes d'information.

5.1.1. La représentation de la tâche d'entrée en relation (EER)

Une tâche se définit en général de manière hiérarchique, elle se décompose en sous-tâches. Les sous-tâches peuvent être définies comme des actions dont l'accomplissement tend à la réalisation de la tâche. Les sous-tâches sont généralement exprimées à partir d'une forme verbale que ce soit le verbe à l'infinitif (*accueillir, demander, donner, …*) ou une nominalisation déverbale (*vérification*).

Après avoir listé les sous-tâches, dans un souci de lisibilité, nous les avons regroupées en différents pôles, en fonction des différents types d'actions sous-jacentes. Nous avons ensuite relevé, pour chacune des sous-tâches, le nombre d'occurrences énoncées par catégorie de sujets.

Les sous-tâches énoncées par les interviewés peuvent être regroupées autour de trois pôles principaux :

- L'*interaction avec le client* (107 énoncés de sous-tâches) : ensemble des actions concernant davantage l'interaction avec le client.
- Le *respect de la législation* (123 énoncés de sous-tâches) : ensemble des actions dont l'objectif principal est le respect de la législation.

- Les *aspects informatiques* (17 énoncés de sous-tâches) : ensemble des actions portant sur les aspects informatiques.

Deux pôles les plus importants (en nombre) peuvent encore être affinés :

- L'interaction avec le client peut se décomposer en trois sous-pôles :
  - *Généralités* (40 sous-tâches) : ensemble d'actions générales gérant la relation avec le client.
  - *Informations précises sur le client* (36 sous-tâches) : ensemble d'actions visant à obtenir des informations précises sur le client.
  - *Informations « marketing »* (31 sous-tâches) : ensemble des actions qui sont davantage orientées commercial.
- Le respect de la législation peut également se décomposer en trois sous-pôles :
  - *Concernant le client* (39 sous-tâches) : ensemble des actions permettant d'assurer la « fiabilité » du client.
  - *Concernant les pièces justificatives* (56 sous-tâches) : ensemble des actions visant à obtenir les pièces justificatives nécessaires à la constitution du dossier et à s'assurer de la validité de ces documents.
  - *Concernant l'ouverture de compte* (28 sous-tâches) : toutes les autres actions nécessaires pour pouvoir réaliser l'ouverture de compte dans les règles.

Nous insistons sur le fait que ces regroupements de sous-tâches ont été réalisés *a posteriori*, pour faciliter la lecture et l'interprétation des résultats concernant les sous-tâches. Ils résultent d'une interprétation de notre part, basée sur les propos tenus par les interviewés au cours des entretiens. Si cette répartition des sous-tâches devait être reprise pour une implémentation quelconque, il conviendrait de réaliser une expérimentation pour la valider.

Le deuxième pôle « respect de la législation » rassemble un peu plus d'occurrences (123) que le pôle « interaction avec le client » (107), mais la différence n'est pas flagrante. Le pôle « saisir les données informatiques » en revanche, recouvre le moins de sous-tâches.
Nous nous sommes intéressés au nombre d'occurrences de chaque sous-tâche en fonction de la catégorie d'utilisateurs. Le nombre d'interviewés par catégorie n'étant pas équivalent, il semble plus pertinent de considérer la moyenne du nombre d'occurrences (cf. tableau 3).

Moyennes et % par Pôles	**Comm. novices**	**Comm. experts**	**Contrôleurs novices**	**Contr. experts**	**Animateurs commerciaux novices**	**Anim. Com. experts**
Interaction client	9,50	6,67	1,00	2,80	11,00	5,75
	48,72 %	**40,82 %**	**11,11 %**	**26,92 %**	**57,89 %**	**57,50 %**
Respect de la législation	9,00	8,67	8,00	5,80	7,00	4,25
	46,15 %	**53,06 %**	**88,89 %**	**55,77 %**	**36,84 %**	**42,50 %**
Aspects informatiques	1,00	1,00	0,00	1,80	1,00	0,00
	5,13 %	**6,12 %**	**0,00 %**	**17,31 %**	**5,26 %**	**0,00 %**
Total du nombre d'occurrences	19,50	16,33	9,00	10,40	19,00	10,00
	100 %	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**

Tableau 3 : Moyennes et pourcentages du nombre d'occurrences des ST mentionnées.

Dans l'ensemble, les commerciaux s'avèrent, en moyenne, plus prolixes que les contrôleurs sur la décomposition de la tâche d'EER en sous-tâches. Ce qui n'est par

ailleurs pas très surprenant dans la mesure où ce sont eux qui ont en charge la plus grande partie de la procédure.

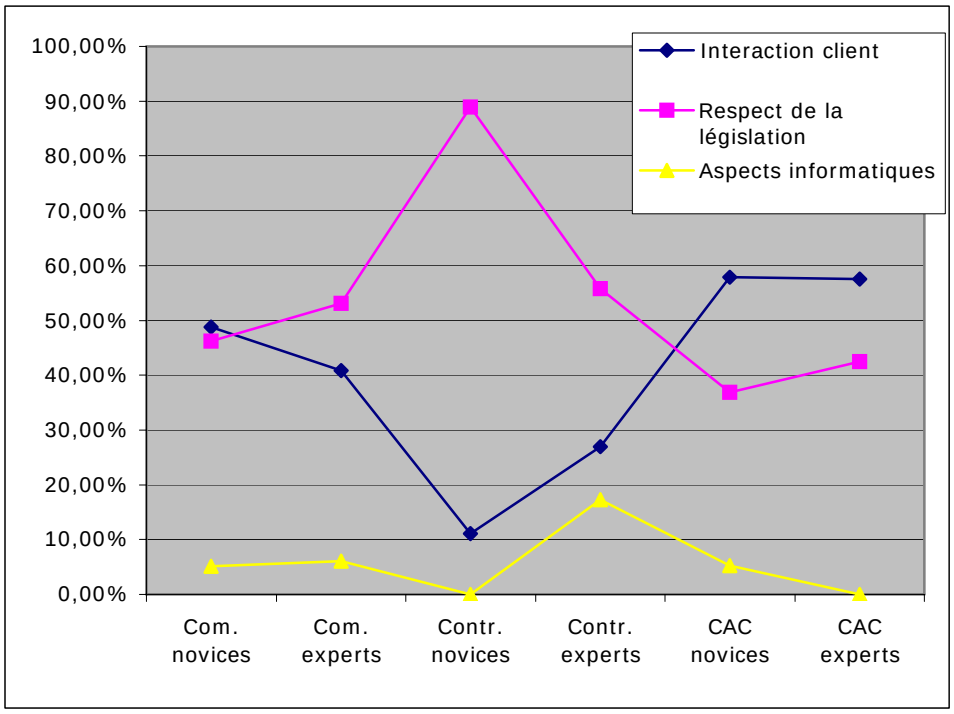


Figure 1 : Moyennes et pourcentages du nombre d'occurrences de sous-tâches mentionnées par groupe d'interviewés.

La figure ci-dessus montre clairement que les contrôleurs citent davantage de sous-tâches liées au respect de la législation qu'ils n'en citent sur l'aspect interaction avec le client. En revanche, contrairement à ce que pensent les contrôleurs, cette distinction n'est pas aussi marquée pour les commerciaux : ces derniers citent quasiment autant de sous-tâches concernant l'aspect interaction avec le client, que de sous-tâches relatives au respect de la législation. La distinction est plus marquée pour les animateurs commerciaux qui eux, citent plus de sous-tâches sur l'interaction client que sur le respect de la législation.

Sans entrer plus avant dans les détails puisque l'analyse de la tâche d'EER n'est pas l'objectif premier de l'étude, il semble que les interviewés, quel que soit leur métier et leur niveau d'expertise, ont une conception semblable de la tâche et des sous-tâches qui la composent. Ainsi, tous citent des activités en lien avec les trois grands pôles identifiés. Ceci étant, chaque métier oriente plus précisément sa conception de la tâche par rapport à sa propre activité : le commercial et l'animateur commercial accorderont une plus grande importance aux clients et aux produits et le contrôleur aux vérifications et au respect de la loi.

5.1.2. Le vocabulaire caractérisant la tâche d'EER

Nous avons étudié le vocabulaire associé à la tâche d'EER et nous pouvons tirer de l'analyse de ce vocabulaire les remarques suivantes :
C'est un vocabulaire extrêmement précis et les sujets sont descendus à un niveau très fin : plus de 230 termes pour caractériser cette seule tâche.

> Exemples : *enfants mineurs d'étrangers résidant en France, conditions de validité des justificatifs.*

Ce sont à la fois des termes simples et des mots composés qui ont été utilisés par les sujets.

Exemples : *client, justificatif, gens du voyage, hébergement en foyer, …*

Même si l'EER est une tâche clairement délimitée et si le domaine bancaire est, comme tous les domaines spécialisés, relativement fermé et utilisant un vocabulaire monosémique, les sujets peuvent employer des termes différents pour recouvrir une même notion.

Exemples : *situation professionnelle et activité professionnelle, vérification et contrôle, saisie des données et saisie des informations.*

L'étude du vocabulaire propre à l'EER montre qu'un certain nombre de pôles se dégagent, autour desquels les termes cités peuvent être regroupés. Les pôles suivants ont été identifiés :

- connaissance du client (accueil, demande de justificatifs,…),
- légalité des opérations (contrôles et vérifications),
- produits et services bancaires,
- fonctionnement des comptes (types de comptes, moyens de paiement),
- procédure informatique d'ouverture de compte.

Tous les interviewés, quel que soient leur métier et leur niveau d'expertise, ont cité du vocabulaire lié à chacun de ces pôles. Des différences ont cependant été notées selon les types de sujets.

- les commerciaux ont été plus précis sur le vocabulaire lié aux clients et aux produits ;
- les contrôleurs ont cité plus de termes sur le pôle « légalité des opérations » ;
- les animateurs commerciaux ont cité plus de termes sur les clients que sur la légalité des opérations ;
- les commerciaux novices ont été plus précis, quel que soit le pôle concerné, que les commerciaux experts ;
- à l'inverse, chez les contrôleurs, ce sont les experts qui ont été le plus précis en citant le plus de termes.

5.1.3. Les demandes d'information dans le cadre de l'EER

Les interviewés ont été amenés à donner des exemples de manques d'information qui peuvent apparaître lors de la réalisation de la tâche d'EER. Ces manques d'information ont été exprimés sous la forme de questions ou de thèmes

La liste des demandes d'information, constituée à la fois de questions réelles et de thèmes susceptibles de devenir des demandes d'information, met en évidence plusieurs points :

- L'importance des besoins d'information dans l'activité quotidienne des interviewés et donc la nécessité pour eux de disposer de sources d'information leur permettant de trouver rapidement des réponses à leurs questions.
- La variété des besoins d'information susceptibles de poser problème dans la pratique de leur activité.
- Le fait enfin que, même pour une tâche qui semble clairement délimitée et même routinière de l'avis de certains sujets, les interviewés sont amenés à s'interroger sur des points non directement liés à la tâche qu'ils ont à effectuer. Par exemple, ils s'interrogent sur les produits, les assurances, la fiscalité, …

Ce constat montre l'importance de l'information et de la recherche d'information dans l'activité des collaborateurs de la banque, et ce quelle que soit leur fonction.

5.2. Résultats de l'expérience 2

On a demandé aux sujets de trier un ensemble de documents. Chaque document représente un extrait d'un document normatif de l'entreprise.
Chaque sujet devait regrouper les documents qui, selon lui, constituent un ensemble homogène. Ensuite, il devait expliquer la différence entre les groupes de documents et devait donner un titre à chaque groupe de documents.
Nous avons ainsi recueilli pour chaque sujet le nombre de groupes et les dénominations de chaque groupe Nous avons étudié ces groupes et leurs dénominations selon les catégories professionnelles.

5.2.1. Les groupes de documents

Les 9 sujets ont produit un total de 47 groupes de documents. À ces 47 groupes de documents correspondent 42 libellés différents. Seuls deux libellés reviennent : Mineurs (3 sujets), Ouverture de compte (2 sujets). On note donc une très grande variabilité dans le choix des titres affectés aux groupes de documents. Il semble que l'expertise et la catégorie métier introduisent des variations à la fois dans le regroupement des documents et dans les choix des libellés pour les groupes.

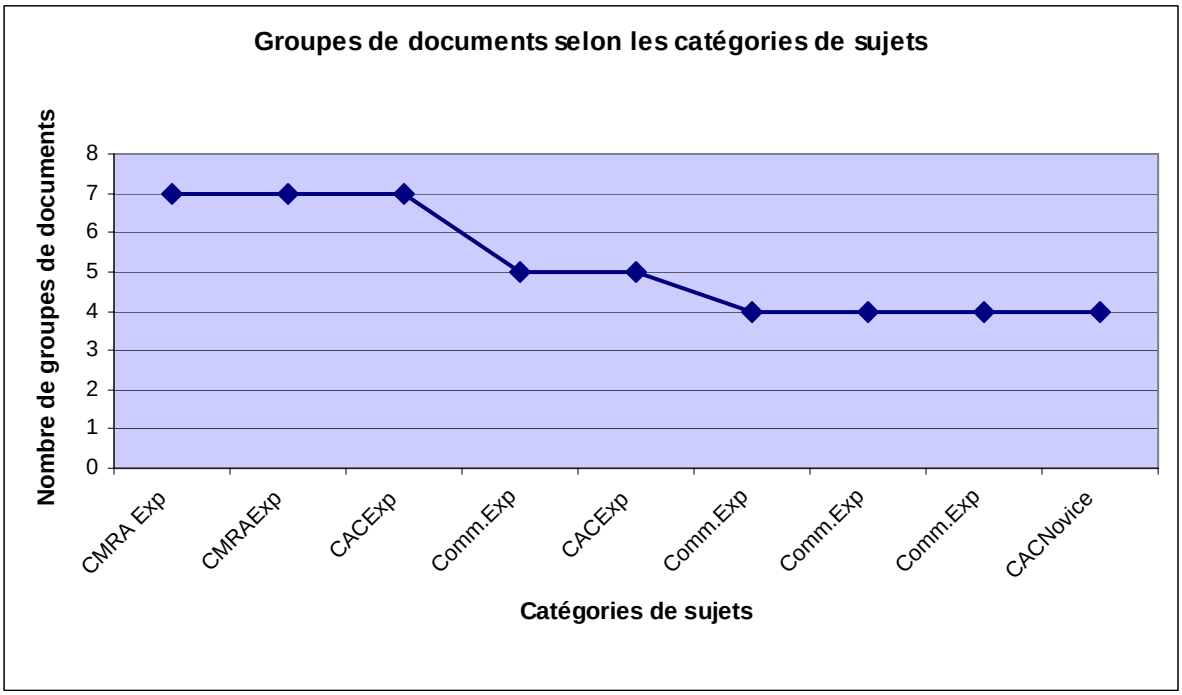


Figure 2: Groupes de documents produits selon les catégories de sujets.

La figure ci-dessus montre que les sujets les plus experts tendent à sérier davantage les documents : ils tendent à faire plus de groupes de documents que les autres sujets. Les animateurs commerciaux experts et les contrôleurs se détachent et il semblerait également que les contrôleurs varient moins que les autres catégories.

Parmi les documents proposés, certains sont plus souvent regroupés ensemble quels que soient le métier et l'expertise du sujet (par exemple, ceux ayant trait aux ouvertures de compte courantes).
Les regroupements opérés par les contrôleurs semblent moins dispersés que ceux opérés par les autres catégories de sujets. Ainsi, les contrôleurs tendent à être plus précis dans la composition des groupes de documents.

Exemple : plusieurs extraits traitaient des types de comptes et d'ouvertures de compte en général. Les contrôleurs tendent à faire des catégories précises :

- un groupe de documents sur le cas général d'ouverture de compte,
- un groupe pour chaque cas particulier d'ouverture de compte,
- un autre groupe sur les pièces à recueillir,
- un autre groupe sur le cas des mineurs.

5.2.2. Les libellés attribués aux groupes de documents

Les noms donnés aux groupes par les sujets sont relativement variables, ce qui est cohérent avec le type de résultats obtenus avec cette technique expérimentale.
Cependant, les contrôleurs ont tendance, pour chaque cas, à choisir des titres précis tels que :

- Comptes spécifiques : PMU, Française des jeux, agents d'assurances ;
- Ouverture de compte Assurances, Agent général ;
- Ouverture de compte pour les majeurs capables ;
- Ouverture de compte Assurances, Agent général ;
- Ouverture de compte PMU, française des jeux ;
- Comptes en indivision ;
- Comptes rémunérés ;
- Mineurs : définition, capacité, pouvoirs, vérification.

À l'intérieur du groupe des contrôleurs, on note relativement peu de variations inter-sujets dans le dans la composition des groupes de documents comme dans le choix des libellés : ainsi : par exemple, pour le même regroupement de documents, on pourra trouver les libellés « Indivision » ou « Compte en indivision ».

Pour les autres catégories professionnelles, la composition des groupes de documents est moins homogène et de ce fait, les libellés sont plus variables :

- Effectivement, les commerciaux choisissent des libellés plus clairs ou plus adaptés aux produits commerciaux comme « ouverture de comptes épargnes et prêts », « Emprunts et crédits ».
- Les commerciaux tendent à donner des libellés plus vagues : Informations générales Banque, Étranger, Juridique, Ouverture de Compte, Règles de l'entrée en relation.
- Les contrôleurs choisissent « :Mineurs : définition, capacité, pouvoirs, vérification » alors que les autres catégories choisissent majoritairement un libellé plus vague tel que « Mineurs ».

6. Discussion et perspectives

6.1. Remarques sur les sujets : de la difficulté à définir l'expertise professionnelle

Les résultats obtenus doivent être nuancés : en effet, l'équilibre entre sujets experts et novices n'a pas pu être respecté dans toutes les étapes expérimentales car les sujets qui nous ont été proposés sont majoritairement des experts. Par exemple dans la catégorie des contrôleurs, certains sujets présentés comme novice, occupent cette fonction depuis un an, mais ils ont été chargés de contrôle pendant 5 ans auparavant. Il est de ce fait difficile de les considérer comme étant de « réels » novices, leur fonctionnement pour la tâche considérée étant plutôt celui d'un expert. Ce cas de figure s'est également retrouvé pour la catégorie des animateurs commerciaux.
De plus**,** si les catégories « commerciaux » et « contrôleurs » semblent homogènes, il n'en est pas de même de la catégorie des animateurs commerciaux. En effet, dans cette

catégorie, nous avons noté des différences dans les fonctions assumées réellement par les sujets (animation d'équipe commerciale, directeur d'entité, directeur d'agence,…). Ceci explique sans doute les résultats plus ambigus dans ces catégories.

6.2. Sur la variété du vocabulaire employé pour décrire une tâche courante

L'étude du vocabulaire recueilli autour d'une tâche très courante dans les entreprises bancaires, montre la grande diversité des dénominations. Une telle remarque avait été déjà faite dans des études précédentes telles que [FURNAS, 1983 ; FURNAS, 1987 ; GOMEZ, 1988 ; SARACEVIC 1988 ; BATES 1998].
Compte tenu du fait qu'on se trouve ici dans un cadre professionnel, on s'attendait à rencontrer un lexique plus restreint. Or, il se trouve que le vocabulaire recueilli comporte à la fois des termes de métiers mais aussi des éléments appartenant à langue courante.

En revanche, si on se limite à la partie « professionnelle » de ce vocabulaire, on remarque que métiers et niveau d'expertise introduisent des différences, ainsi on note que :

- C'est un vocabulaire précis, dont les termes se regroupent autour de quelques pôles thématiques partagés par tous les sujets quel que soit leur métier ou niveau d'expertise.
- Selon les métiers, il existe des différences dans le choix des termes pour donner une représentation de la tâche : selon leur activité, les sujets accentuent plus sur un pôle thématique (client) ou sur un autre (légalité).
- Le vocabulaire utilisé par les contrôleurs est un vocabulaire technique plus précis alors que les commerciaux, pour désigner les mêmes objets, auront tendance à employer un langage courant.
- De ce fait, on peut comprendre que le vocabulaire propre aux contrôleurs est d'avantage en adéquation avec le vocabulaire employé dans les documents normatifs.

6.3. Améliorer les procédures de recherche

Améliorer l'accès aux documents passe, en partie, par l'amélioration des modes de recherche et d'indexation des documents en tenant compte des caractéristiques des usagers du système d'information. Le problème principal est ici d'améliorer l'accès aux documents pertinents, dans le cadre d'un système de recherche d'information en langue naturelle. En effet, nous avons vu que seuls les employés les plus experts peuvent accéder aux documents à partir du plan de classement actuel. Par ailleurs, nous avons vu que les regroupements d'informations varient non seulement avec l'expertise mais aussi les différents métiers. Il semble peu réaliste d'imaginer la mise en place rapide d'un nouveau plan de classement qui réponde aux critères des « différents métiers ».

Nous avons vu également que d'une manière générale, les usagers délaissent le thésaurus spécialisé existant dans l'entreprise. Conformément à d'autres analyses [WELLISCH 1996], [WALCHODER, 2006], ils tendent à préférer leur propre vocabulaire pour une recherche en langage naturel, malgré tous les problèmes liés à ce type d'interrogation. Sans doute, le thesaurus existant pourrait être enrichi par l'intégration du vocabulaire utilisateur mais, serait-il pour autant plus utilisé ? Sans doute, le thesaurus existant pourrait être enrichi par l'intégration du vocabulaire utilisateur mais, serait-il pour autant plus utilisé ?

Une voie différente serait celle de l'intégration d'une ontologie qui répondrait à la nécessité d'un intermédiaire pour l'interrogation en langue naturelle, tout en évitant les écueils de la mise en œuvre d'un langage documentaire traditionnel. Cette ontologie devra englober des concepts issus des documents et des concepts provenant des usagers

des documents. Dans le cas des documents normatifs étudiés, des méthodes de traitement automatique de la langue et d'extraction d'information peuvent convenir pour l'extraction des concepts et des relations qu'ils entretiennent.
Pour compléter ce premier inventaire, il est nécessaire de repérer les tâches caractéristiques de chaque « métier » et d'extraire le vocabulaire particulier afférant à chacune de d'elles. Compte tenu de la richesse du vocabulaire recueilli via des protocoles oraux, il serait utile d'étendre cette technique de recueil à d'autres activités propres au secteur bancaire. La difficulté viendra ensuite de l'exploitation automatique de ce type de corpus, du fait des particularités linguistiques qu'il peut présenter. Les méthodes habituelles de traitement de la langue naturelle ne pourront pas s'appliquer à de tels corpus. Néanmoins, d'autres méthodes ont été proposées pour exploiter automatiquement des notes de communications orales [EVEN, 2005] ou des corpus d'entretiens [EVEN, 2002]. Dans ce dernier cas, un corpus clients-employés dans le secteur bancaire a été utilisé. Les corpus oraux doivent en premier lieu passer par une phase de normalisation avant de construire une ontologie.

7. Conclusions

Malgré la spécificité du domaine professionnel, le vocabulaire du secteur de la banque n'est pas aussi homogène qu'on pouvait l'espérer. Les variations lexicales semblent liées à plusieurs facteurs :

- A l'existence d'une part de lexique non professionnel, relevant de la langue générale, qui est normalement sujette à beaucoup de variations,
- au vocabulaire strictement professionnel : celui-ci étant lié à la compétence des utilisateurs, mais aussi aux fonctions qu'ils assurent.

On voit bien que ces variations de vocabulaire posent des problèmes d'une part dans l'accès à l'information et d'autre part dans l'exploitation et la compréhension des informations retrouvées.
L'amélioration de l'accès à l'information passe donc par la prise en compte des utilisateurs visés. Cette prise en compte repose sur leur niveau d'expertise dans leur fonction mais aussi sur les activités qu'ils ont l'habitude de prendre en charge quotidiennement. .

Plusieurs solutions peuvent être envisagées pour remédier à cette difficulté d'accès à l'information :

- Le premier type de réponse concerne la conception des documents. Il est effectivement envisageable de réécrire les documents d'entreprise en fonction des lecteurs visés. Ainsi, un même document pourrait être conçu de différentes manières en fonction de la catégorie d'employés visée, en tenant compte du vocabulaire propre à cette classe d'utilisateurs. Cette solution suppose que l'on mette en place des dispositifs de guidage inter-documents pour permettre aux utilisateurs l'accès à tous les documents relatifs à une question.
- Il est également possible de proposer un même document pour tous les utilisateurs mais avec des modes d'accès différents, adaptés là aussi à chaque catégorie d'employés.
- Enfin, l'intégration, dans les langages d'indexation et de recherche, du vocabulaire ou des vocabulaires utilisés dans l'entreprise devrait concourir à améliorer l'accès aux documents d'entreprise.

Références

[BATES, 1998] Bates M..J., Indexing and access for digital libraries and the Internet : Human, database and domain factors. Journal of the American Society for Information Science, 49, vol. 13, 1185–1205, 1998.

[BISSERET, 1999] Bisseret A., Sebilotte S., Falzon P., Techniques pratiques pour l'étude des activités expertes. Toulouse, ed. Octares, 1999.

[BRUCE, 1994] Bruce H.W., A cognitive view of the situational dynamism of user-centered relevance estimation , *Journal of the American Society for Information Science*, vol. 45, p. 142-148, 1994.

[EVEN, 2002] Even F., Enguehard C. Extraction d'informations à partir de corpus dégradés. Actes de la Neuvième Conférence Nationale sur le Traitement Automatique des Langues Naturelles (TALN 2002), 2002, Volume 1, page 105-114. Disponible sur http://www.loria.fr/projets/JEP-TALN/actes/TALN/articles/TALN09.pdf (Consulté le 10 décembre 2006).

[EVEN, 2005] Even F. Extraction d'information et modélisation de connaissances à partir de notes de communication orale. Thèse en Informatique, Université de Nantes, 252 p, 2005.

[FURNAS, 1983] Furnas G.W., Landauer T.K., Gomez L.M., Dumais S.T. Human factors and behavioral science : statistical semantics, analysis of the potential performance of key-word information system, *The Bell system technical journal*, vol. 62, n°6, aout-juillet. 1983.

[FURNAS, 1987] Furnas, G., Landauer, T., Gomez, L., Dumais, S.T. The vocabulary problem in human–system communication. Communications of the ACM, 30, vol. 964–971, 1987.

[GOMEZ, 1988] Gomez L., Lochbaum, C.C., Landauer, T.K. All the right words: Finding what you want as a function of the richness of indexing vocabulary. *Journal of the American Society for Information Science*, 41, vol. 8, 547–559, 1988.

[HARTER, 1992] Harter, S.P., Psychological relevance and information science , *Journal of the American Society for Information Science*, vol. 43, p. 602-615, 1992.

[LE COADIC, 1992] Le Coadic Y.F., *Usages et usagers de l'information*, Paris : Nathan, 1997.

[MARCHIONINI, 1991] Marchionini G., Meadows C., Dwiggins S., Lin X., Wang J., Yuan W. A study of user interaction with information retrieval interfaces : Progress report, *Canadian Journal of Information Science*, vol. 16, p. 42-59, 1991.

[PEILING, 1998] Peiling W., Sorgel D., A cognitive model of document use during a research project, Study 1, Document selection, *Journal of the American Society for Information Science*, vol. 49(2), p. 115-125, 1998.

[SARACEVIC, 1988] Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. A study of information seeking and retrieving: I. Background and methodology. *JASIS*, 39, vol.3, 161–176, 1988.

[WACHOLDER, 2006] Wacholder N., Liu L. User preference: a measure of query-term quality. *JASIST*, 57-12, pp.1566-1580, 2006.

[WELLISH, 1996] Wellish M.H. Indexing from A to Z. New York: H.W. Wilson, 1996.